

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
 Art Unit: _____ Phone Number 30 _____ Serial Number: _____
 Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

Point of Contact:
 Jan Delaval
 Librarian-Physical Sciences
 CM1 1E01 Tel: 308-4498

STAFF USE ONLY

Searcher: Jan
 Searcher Phone #: 4446
 Searcher Location: _____
 Date Searcher Picked Up: 1/14
 Date Completed: 1/14
 Searcher Prep & Review Time: _____
 Clerical Prep Time: 15
 Online Time: + 60

Type of Search

NA Sequence (#) _____
 AA Sequence (#) 26
 Structure (#) _____
 Bibliographic _____
 Litigation _____
 Fulltext _____
 Patent Family _____
 Other _____

Vendors and cost where applicable

STN _____
 Dialog _____
 Questel/Orbit _____
 Dr.Link _____
 Lexis/Nexis _____
 Sequence Systems ☒
 WWW/Internet _____
 Other (specify) _____

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:22:30 ; Search time 63.57 Seconds
(without alignments)
9.586 Million cell updates/sec

Title: 09-185908-1A
Perfect score: 36
Sequence: 1 wkxxyxg 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database :
1: PIR_68: *
2: PIR1: *
3: PIR2: *
4: PIR3: *
5: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	86.1	65	2 S23164	light-harvesting p
2	83.3	159	1 QOCV2	aphid transmission
3	83.3	246	2 T32510	hypothetical prote
4	80.6	1180	1 NCECX5	exodeoxyribonuclea
5	80.6	1180	2 G85933	hypothetical prote
6	80.6	1208	2 B82091	exodeoxyribonuclea
7	77.8	278	2 H71119	hypothetical prote
8	77.8	341	2 T48858	acetylpolymine am
9	77.8	424	2 D75350	probable beta-lact
10	77.8	475	2 T02313	endoplasmic reticu
11	77.8	475	2 F86415	hypothetical prote
12	77.8	475	2 B96816	hypothetical prote
13	77.8	488	2 B69415	group II decarboxy
14	77.8	519	2 S45345	Tup1-like enhancer
15	77.8	528	2 G70854	probable sera prot
16	77.8	528	2 T45418	phosphoglycerate d
17	77.8	529	2 T35831	probable D-3-phosp
18	77.8	551	2 T16557	hypothetical prote
19	77.8	766	2 S45344	Tup1-like enhancer
20	77.8	1015	2 S68141	nuclear protein HI
21	77.8	1017	2 I37465	HTR protein - hum
22	77.8	1038	2 JC6027	11k outer membran
23	77.8	4152	2 T31102	filamentous hemagg
24	77.8	4919	2 T31105	hypothetical prote
25	75.0	109	2 D82576	hypothetical prote
26	75.0	116	2 S10864	early E4 11k prote
27	75.0	133	2 A96706	unknown protein, 6
28	75.0	241	2 B83447	hypothetical prote
29	75.0	261	2 C64948	probable membrane

30	27	75.0	261	2 E85798	hypothetical prote
31	27	75.0	348	1 DEBYA	alcohol dehydrogen
32	27	75.0	348	1 DEBYA2	alcohol dehydrogen
33	27	75.0	348	1 T18230	alcohol dehydrogen
34	27	75.0	359	2 T02011	probable cathepsin
35	27	75.0	449	2 T04107	calmodulin-binding
36	27	75.0	493	2 A48457	hexokinase (EC 2.7
37	27	75.0	519	2 S77572	oligopeptide trans
38	27	75.0	523	2 T26696	hypothetical prote
39	27	75.0	531	2 T21576	hypothetical prote
40	27	75.0	537	2 T33084	hypothetical prote
41	27	75.0	538	2 F83354	probable sulfatase
42	27	75.0	542	2 T18912	hypothetical prote
43	27	75.0	590	2 T50195	hypothetical prote
44	27	75.0	605	2 S01066	regulatory protein
45	27	75.0	618	2 T05518	hypothetical prote
46	27	75.0	649	2 T32755	hypothetical prote
47	27	75.0	660	2 A84232	spore cortex synth
48	27	75.0	678	2 T05821	hypothetical prote
49	27	75.0	748	2 A56047	gamma-interferon a
50	27	75.0	851	2 A46160	interferon alpha-1

ALIGNMENTS

RESULT 1
S23164
light-harvesting protein alpha chain - Ectochlorhodospira halochloris
N:Alternate names: antenna pigment protein alpha chain
C:Species: Ectochlorhodospira halochloris
C:Date: 19-Mar-1997 #sequence-revision 19-Mar-1997 #text-change 21-Aug-1998
C:Accession: S23164
R:Wagner-Huber, R.; Brunisholz, R.A.; Bissig, I.; Frank, G.; Suter, F.; Zuber, H.
Eur. J. Biochem. 205, 917-925, 1992
A:title: The primary structure of the antenna polypeptides of Ectochlorhodospira halo
A:reference number: S23164; MUID:92249336
A:Accession: S23164
A:Molecule type: protein
A:Residues: 1-65 <WAG>
C:Superfamily: light-harvesting protein alpha chain
C:Keywords: antenna complex; bacteriochlorophyll; blocked amino end; light-harvesting
F:1/Modified site: N-formylmethionine #status experimental

Query Match 86.1%; Score 31; DB 2; Length 65;
Best Local Similarity 62.5%; Pred. NO. 2.9;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 wkxxyxg 8
DB 58 WKRTSYDG 65

RESULT 2
QOCV2
aphid transmission protein - cauliflower mosaic virus
N:Alternate names: ORF II; ORF2 protein
C:Species: cauliflower mosaic virus
C:Date: 31-Oct-1980 #sequence-revision 31-Oct-1980 #text-change 23-Jul-1999
C:Accession: A90799; JAO010; A93729; A94613; JN0494; A04157
R:Frank, A.; Guillely, H.; Jonard, G.; Richards, K.; Hirth, L.
Cell 21, 285-294, 1980
A:title: Nucleotide sequence of cauliflower mosaic virus DNA.
A:reference number: A90799; MUID:81001865
A:Accession: A90799
A:Molecule type: DNA
A:Residues: 1-159 <FRA>
A:Cross-references: GB:V00141; GB:J02048; NID:958821; PIDN:CAA23457.1; PID:958823
A:Experimental source: strain Strasbourg
R:Modjtabadi, N.; Volovitch, M.; Mazzolini, L.; Yot, P.
FEBS Lett. 181, 223-228, 1985
A:title: Comparison of the predicted secondary structure of aphid transmission factor

A:Reference number: A91334
 A:Accession: J04010
 A:Molecule type: DNA
 A:Residues: 1159 <MOD>
 A:Experimental source: strain PV14
 R:Gardner, R.C.; Howarth, A.J.; Hahn, P.; Brown-Luedi, M.; Shepherd, R.J.; Messing, J.
 Nucleic Acids Res. 9, 2871-2888, 1981
 A:Title: The complete nucleotide sequence of an infectious clone of cauliflower mosaic virus
 A:Reference number: A93729; MUID:82014878
 A:Accession: A93729
 A:Molecule type: DNA
 A:Residues: 1-88, 'N', 90-93, 'R', 95-104, 'V', 106-117, 'N', 119-120, 'N', 123-126, 'D', 128-159 <G
 A:Experimental source: strain CM1841
 R:Guilley, H.
 submitted to the Nucleic Acid sequence Database, October 1982
 A:Reference number: A94613
 A:Accession: A94613
 A:Molecule type: DNA
 A:Residues: 1-50, 'K', 52-101, 'P', 103-126, 'DE', 129-159 <GUL>
 A:Experimental source: strain D/H
 R:Chenault, K.D.; Melcher, U.
 Gene 123, 255-257, 1993
 A:Title: The complete nucleotide sequence of cauliflower mosaic virus isolate BBC.
 A:Reference number: JN0493; MUID:33154593
 A:Accession: JN0494
 A:Molecule type: DNA
 A:Residues: 1, 'R', 3-117, 'N', 119-126, 'D', 128-137, 'K', 139-159 <CHE>
 A:Cross-references: GB:W90542; NID:9678542; PIDN:AA62372.1; PID:929182
 A:Experimental source: isolate BBC
 C:Superfamily: cauliflower mosaic virus aphid transmission protein

Query Match 83.3%; Score 30; DB 1; Length 159;
 Best Local Similarity 62.5%; Pred. No. 12;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsyxg 8
 Db 61 WKINSTYG 68

RESULT 3
 T32510
 hypothetical protein C44B12.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 04-Mar-2000
 C:Accession: T32510
 R:Pin-Mollam, A.
 submitted to the EMBL Data Library, December 1997
 A:Description: The sequence of C. elegans cosmid C44B12.
 A:Reference number: Z21183
 A:Accession: T32510
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-246 <PIN>
 A:Cross-references: EMBL:AR036692; PIDN:AA88326.1; GSPDB:GN00022; CESP:C44B12.3
 A:Experimental source: strain Bristol N2; clone C44B12
 C:Genetics:
 A:Gene: CESP:C44B12.3
 A:Map position: 4
 A:Introns: 73/3; 112/3; 137/1; 202/1
 C:Superfamily: Caenorhabditis elegans hypothetical protein C44B12.3

Query Match 83.3%; Score 30; DB 2; Length 246;
 Best Local Similarity 62.5%; Pred. No. 18;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsyxg 8
 Db 87 WKOKSYDG 94

RESULT 4
 NCECX5
 exodeoxyribonuclease V (EC 3.1.11.5) 135K chain - Escherichia coli
 N:Alternate names: exonuclease 135K polypeptide; recBC Dnase 135K polypeptide
 C:Species: Escherichia coli
 C:Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 19-Jan-2001
 C:Accession: A25532; E65064
 R:Finch, P.W.; Storey, A.; Chapman, K.E.; Brown, K.; Hickson, I.D.; Emerson, P.T.
 Nucleic Acids Res. 14, 8573-8582, 1986
 A:Title: Complete nucleotide sequence of the Escherichia coli recB gene.
 A:Reference number: A25532; MUID:87066729
 A:Accession: A25532
 A:Molecule type: DNA
 A:Residues: 1-1180 <FIN>
 A:Cross-references: GB:X04581; NID:942680; PIDN:CAA28250.1; PID:942682
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
 A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617
 A:Accession: E65064
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1180 <BLAT>
 A:Cross-references: GB:AE000365; GB:U00096; NID:92367163; PIDN:AA075859.1; PID:917891
 A:Experimental source: strain K-12, substrain MG1655
 C:Comment: This enzyme is required for efficient DNA repair; it catalyzes the unwind
 11 of these activities require concomitant hydrolysis of ATP.
 C:Genetics:
 A:Gene: recB
 A:Map position: 61 min
 C:Superfamily: exodeoxyribonuclease V 135K chain
 C:Keywords: ATP; DNA repair; hydrolase; nucleotide binding; P-loop
 F:23-30/Region: nucleotide-binding motif A (P-loop)

Query Match 80.6%; Score 29; DB 1; Length 1180;
 Best Local Similarity 50.0%; Pred. No. 1,5e+02;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsyxg 8
 Db 901 WRVTSYSG 908

RESULT 5
 G85933
 hypothetical protein recB [imported] - Escherichia coli (strain O157:H7)
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
 C:Accession: G85933
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
 Jiller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: G85933
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1180 <STO>
 A:Cross-references: GB:AE005174; NID:912517302; PIDN:AA657931.1; GSPDB:GN00145; UMG:
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: recB
 C:Superfamily: exodeoxyribonuclease V 135K chain

Query Match 80.6%; Score 29; DB 2; Length 1180;
 Best Local Similarity 50.0%; Pred. No. 1,5e+02;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsyxg 8
 Db 11

Db 901 WRVMSYSG 908

RESULT 6
B82091
exodeoxyribonuclease V, 135 kDa chain VC2320 [imported] - Vibrio cholerae (strain N16961)
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: B82091
R:Heidelberg, J.F.; Elsen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Karlsson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Drygala, I.; Sellers, F.
L.; R.R.; McKalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82055; MUID:20406833
A:Accession: B82091
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1208 <HEI2>
A:Cross-references: GB:AE004303; GB:AE003852; NID:g9656890; PIDN:AAF95464.1; GSPDB:GN001
C:Genetics:
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
A:Gene: VC2320
A:Map position: 1
C:Superfamily: exodeoxyribonuclease V 135k chain

Query Match 80.6%; Score 29; DB 2; Length 1208;
Best Local Similarity 50.0%; Pred. No. 1.5e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxyxg 8
|||
Db 906 WRVMSYSG 913

RESULT 7
H71119
hypothetical protein PH0727 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 05-Nov-1999
C:Accession: H71119
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekir
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A:Reference number: A71000; MUID:98344137
A:Accession: H71119
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-278 <KAM>
A:Cross-references: GB:AF000003; NID:g3236130; PIDN:BA29818.1; PID:d1030761; PID:g32571
A:Experimental source: strain OT3
A:Note: This accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH0727

Query Match 77.8%; Score 28; DB 2; Length 278;
Best Local Similarity 50.0%; Pred. NO. 59;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxyxg 8
|||
Db 99 WRVMSYSG 106

RESULT 8
T48858
acetylpolymaline aminohydroxylase [validated] - Mycoplasma ramosa
C:Species: Mycoplasma ramosa
C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C:Accession: T48858

R:Sakurada, K.; Chata, T.; Fujishiro, K.; Hasegawa, M.; Aisake, K.
J. Bacteriol. 178, 5761-5766, 1996
A:Title: Acetylpolymaline amido-hydroxylase from Mycoplasma ramosa: gene cloning and chara
A:Reference number: 224559
A:Accession: T48858
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-341 <SAK>
A:Cross-references: EMBL:D10463; PIDN:BAAD1256.1
A:Experimental source: ATCC 49678; strain FERM BP-1845
C:Genetics:
A:Note: apna
C:Function:
A:Description: involved in degradation of acetylpolymalines [validated; MUID:96422009]

Query Match 77.8%; Score 28; DB 2; Length 341;
Best Local Similarity 50.0%; Pred. NO. 72;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wkxxyxg 8
|||
Db 78 WKAAGYKG 85

RESULT 9
D75330
probable beta-lactamase - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: D75330
R:White, O.; Elsen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Ulterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A5250; MUID:20036896
A:Accession: D75330
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-424 <WHI2>
A:Cross-references: GB:AE002036; GB:AE000513; NID:g6459766; PIDN:AAF11537.1; PID:g645
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1985
A:Map position: 1

Query Match 77.8%; Score 28; DB 2; Length 424;
Best Local Similarity 50.0%; Pred. NO. 90;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxyxg 8
|||
Db 339 WKLWYSG 346

RESULT 10
T02313
endoplasmic reticulum insertion protein P13P17.9 - Arabidopsis thaliana
N:Alternate names: hypothetical protein Atg34250
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C:Accession: T02313; C84754
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K
submitted to the EMBL data library, July 1998
A:Description: Arabidopsis thaliana chromosome II BAC P13P17 genomic sequence.
A:Accession: T02313
A:Reference number: Z14657
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-475 <ROU>
A:Cross-references: EMBL:AC004481; NID:g3337347; PIDN:AAC27401.1; PID:g3337356

A:Experimental source: cultivar Columbia
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffatt, K.S.; Cronin, L.A.; Shen, M.; Vankken, S.E.; Unayam, L.; Tallon,
 D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487
 A:Accession: C84754
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-475 <STO>
 A:Cross-references: GB:AE002093; NID:g3337356; PIDN:AAC27401.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: ATg34230; F13P17.9
 A:Map position: 2
 A:Introns: 115/2; 191/2; 309/3; 351/2; 396/3
 C:Superfamily: Yeast SSH1 protein
 C:Keywords: endoplasmic reticulum; protein transport; transmembrane protein

Query Match 77.8%; Score 28; DB 2; Length 475;
 Best Local Similarity 50.0%; Pred. No. 1e+02;
 Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wkxxyxyg 8
 || |
 Db 327 WKSEYSG 334

RESULT 11
 F86415
 Hypothetical protein AAF8109.1 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: F86415
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 Ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzalla,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: F86415
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-475 <STO>
 A:Cross-references: GB:AE005172; NID:g9502410; PIDN:AAF8109.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1
 C:Superfamily: yeast SSH1 protein

Query Match 77.8%; Score 28; DB 2; Length 475;
 Best Local Similarity 50.0%; Pred. No. 1e+02;
 Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wkxxyxyg 8
 || |
 Db 327 WKSEYSG 334

RESULT 12
 B96816
 Hypothetical protein F9K20.24 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
 C:Accession: B96816
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
 C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzalla
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: B96816
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-475 <STO>
 A:Cross-references: GB:AE005173; NID:g3834321; PIDN:AAC83037.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: F9K20.24
 A:Map position: 1
 C:Superfamily: yeast SSH1 protein

Query Match 77.8%; Score 28; DB 2; Length 475;
 Best Local Similarity 50.0%; Pred. No. 1e+02;
 Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wkxxyxyg 8
 || |
 Db 327 WKSEYSG 334

RESULT 13
 B69415
 Group II decarboxylase homolog - Archaeoglobus fulgidus
 C:Species: Archaeoglobus fulgidus
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Jun-1999
 C:Accession: B69415
 R:Kleek, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
 .; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.
 Gload, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
 Nature 390, 364-370, 1997
 A:Authors: Urtreback, T.; Cotton, M.D.; Spriggs, T.; Arlach, P.; Kaine, B.F.; Sykes,
 Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
 A:Reference number: A69250; MUID:98049343
 A:Accession: B69415
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-488 <KLE>
 A:Cross-references: GB:AE001012; GB:AE000782; NID:g2689335; PIDN:AAB89922.1; PID:g264
 C:Superfamily: Escherichia coli glutamate decarboxylase

Query Match 77.8%; Score 28; DB 2; Length 488;
 Best Local Similarity 50.0%; Pred. No. 1e+02;
 Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wkxxyxyg 8
 || |
 Db 166 WKSAEYLG 173

RESULT 14
 S45345
 Tup1-like enhancer - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 20-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 26-May-2000
 C:Accession: S45345
 R:Halford, S.; Wade, R.; Roberts, C.; Daw, S.C.M.; Whiting, J.A.; O'Donnell, H.; Dun
 J.; Cross, I.; Burn, J.; Scambler, P.J.
 Hum. Mol. Genet. 2, 2099-2107, 1993
 A:Title: Isolation of a putative transcriptional regulator from the region of 22q11 d
 A:Reference number: S45345; MUID:94154665
 A:Accession: S45345
 A:Status: preliminary

A:Molecule type: mRNA
A:Residues: 1-519 <HAI>
A:Cross-references: EMBL:X75295; NID:g434992; PIDs:CAA53043.1; PID:g434993
C:Superfamily: unassigned WD repeat proteins; WD repeat homology <WD1>
F:22-55/Domain: WD repeat homology <WD1>
F:126-159/Domain: WD repeat homology <WD2>

Query Match 77.8%; Score 28; DB 2; Length 519;
Best Local Similarity 50.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxyxg 8
DB 53 WKRAVYIG 60

RESULT 15
G70854
Probable *serA* protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: G70854
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: AF0500; MUID:98295987
A:Accession: G70854
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-528 <COL>
A:Cross-references: GB:AL021287; GB:AL123456; NID:g3261508; PIDs:CA16081.1; PID:g279159
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: *serA*
C:Superfamily: Bacillus phosphoglycerate dehydrogenase

Query Match 77.8%; Score 28; DB 2; Length 528;
Best Local Similarity 50.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxyxg 8
DB 129 WKRSFSG 136

RESULT 16
T45418
Phosphoglycerate dehydrogenase [Imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 18-Feb-2000
C:Accession: T45418
R:Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
A:Submitted to the EMBL Data Library, September 1997
A:Reference number: Z16918
A:Accession: T45418
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-528 <PAR>
A:Cross-references: EMBL:Z99263; PIDs:CA16440.1
A:Experimental source: cosmid B637
C:Genetics:
A:Note: *serA*
C:Superfamily: Bacillus phosphoglycerate dehydrogenase

Query Match 77.8%; Score 28; DB 2; Length 528;
Best Local Similarity 50.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxyxg 8
DB 129 WKRSFSG 136

RESULT 17
T35831
probable D-3-phosphoglycerate dehydrogenase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 21-Jan-2000
C:Accession: T35831
R:Murphy, L.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
A:Submitted to the EMBL Data Library, February 1999
A:Reference number: Z21589
A:Accession: T35831
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-529 <MUR>
A:Cross-references: EMBL:AL035569; PIDs:CA137591.1; GSDB:GN00070; SCOEDB:SC8D9.27
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC8D9.27
C:Superfamily: Bacillus phosphoglycerate dehydrogenase

Query Match 77.8%; Score 28; DB 2; Length 529;
Best Local Similarity 50.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wxxxxxyxg 8
DB 130 WKRSKYTG 137

RESULT 18
T16557
hypothetical protein K04E7.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T16557
R:Nhan, M.
A:Submitted to the EMBL Data Library, October 1995
A:Description: The sequence of C. elegans cosmid K04E7.
A:Reference number: Z18535
A:Accession: T16557
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-551 <NHA>
A:Cross-references: EMBL:U39666; NID:g1049408; PIDs:AA0411; PIDs:AAA80412.1; CESP:K0
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:K04E7.3
A:Introns: 43/3; 81/3; 112/3; 153/1; 191/2; 213/3; 241/2; 281/3; 324/1; 364/1; 393/3;

Query Match 77.8%; Score 28; DB 2; Length 551;
Best Local Similarity 50.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxyxg 8
DB 376 WRNISTYG 383

RESULT 19
S45344
TUP1 like enhancer - human
C:Species: Homo sapiens (man)
C:Date: 20-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 26-May-2000
C:Accession: S45344
R:Halford, S.; Wade, R.; Roberts, C.; Daw, S.C.M.; Whitting, J.A.; O'Donnell, H.; Dun
J.; Cross, I.; Burn, J.; Scambler, P.J.

Hum. Mol. Genet. 2, 2099-2107, 1993
 A:Title: Isolation of a putative transcriptional regulator from the region of 22q11 dele
 A:Reference number: S45344; MUID:94154685
 A:Accession: S45344
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-766 <HAL>
 A:Cross-references: EMBL:X75296; NID:9434982; PIDN:CAA53044.1; PID:9434983
 A:Note: the authors did not translate the codon for residue 201
 C:Superfamily: unassigned WD repeat proteins; WD repeat homology
 F:22-55/Domain: WD repeat homology <WD1>
 F:83-116/Domain: WD repeat homology <WD2>
 F:126-159/Domain: WD repeat homology <WD3>

Query Match 77.8%; Score 28; DB 2; Length 766;
 Best Local Similarity 50.0%; Pred. No. 1.6e+02;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxyxg 8
 11:11
 Db 53 WKRAYIG 60

RESULT 20

S68141
 nuclear protein HIRA - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text-change 26-May-2000
 C:Accession: S68141
 R:Scamps, C.; Lordin, S.; Lamour, V.; Lipinski, M.
 Blochim. Biophys. Acta 1306, 5-8, 1996
 A:Title: The HIR protein family: isolation and characterization of a complete murine cDN
 A:Reference number: S68141; MUID:96201697
 A:Accession: S68141
 A:Status: preliminary; nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-1015 <SCA>
 A:Cross-references: EMBL:X92590; NID:gl135984; PIDN:CAA63334.1; PID:e208369; PID:gl3598
 C:Genetics:
 A:Gene: HIRA
 C:Superfamily: unassigned WD repeat proteins; WD repeat homology
 F:66-99/Domain: WD repeat homology <WD1>
 F:127-160/Domain: WD repeat homology <WD2>
 F:170-203/Domain: WD repeat homology <WD3>

Query Match 77.8%; Score 28; DB 2; Length 1015;
 Best Local Similarity 50.0%; Pred. No. 2.2e+02;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxyxg 8
 11:11
 Db 97 WKRAYIG 104

Search completed: January 14, 2002, 07:37:25
 Job time: 895 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:32:14 : Search time 37.71 Seconds
(without alignments)
7.778 Million cell updates/sec

Title: 09-185908-1A
Perfect score: 36
Sequence: 1 wxxxxyxg 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : SwissProtL39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	88.9	230	1	CUD2_HUMAN
2	31	86.1	65	1	LHA2_ECTHL
3	31	86.1	211	1	CLD1_MOUSE
4	31	86.1	211	1	CLD1_RAT
5	30	83.3	99	1	VAT_CAMV
6	30	83.3	159	1	VAT_CAMV
7	30	83.3	159	1	VAT_CAMV
8	30	83.3	159	1	VAT_CAMV
9	30	83.3	159	1	VAT_CAMV
10	30	83.3	159	1	VAT_CAMV
11	29	80.6	230	1	EX5B_ECOLI
12	29	80.6	1180	1	ES5B_ECOLI
13	28	77.8	191	1	CUD7_RAT
14	28	77.8	211	1	CLD1_HUMAN
15	28	77.8	211	1	CLD7_HUMAN
16	28	77.8	211	1	CLD7_MOUSE
17	28	77.8	215	1	CLDX_BRARE
18	28	77.8	341	1	APRA_MYCRA
19	28	77.8	528	1	SERA_MYCIE
20	28	77.8	528	1	SERA_MYCTU
21	28	77.8	551	1	YVD3_CAEBL
22	28	77.8	1015	1	HIRA_MOUSE
23	28	77.8	1017	1	HIRA_HUMAN
24	28	77.8	1018	1	HIRA_CHICK
25	27	75.0	116	1	E411_ADEI2
26	27	75.0	159	1	VAT_CAMV
27	27	75.0	261	1	ZNUB_ECOLI
28	27	75.0	347	1	ADH1_YEAST
29	27	75.0	347	1	ADH2_YEAST
30	27	75.0	493	1	HXX_PLAFA
31	27	75.0	582	1	NIFA_BRAJA
32	27	75.0	614	1	NTBE_RABIT
33	27	75.0	748	1	SPRA4_HUMAN

34	27	75.0	749	1	STPA4_MOUSE	P42228	mus musculus
35	27	75.0	851	1	STPA2_HUMAN	P52630	homo sapien
36	27	75.0	1450	1	MPSE_CHICK	002173	gallus gall
37	26	72.2	83	1	CYC6_PAVLU	P00107	pavlova lut
38	26	72.2	140	1	RS19_SULSO	09uxa3	sulfolobus
39	26	72.2	391	1	V534_MERTJA	057954	methanococ
40	26	72.2	404	1	FD3C_BRANA	P46618	brassica na
41	26	72.2	411	1	INTR_ECOLI	P76056	escherichia
42	26	72.2	420	1	MSME_STRMU	000749	streptococc
43	26	72.2	435	1	AM3D_ORYSA	P27933	oryza sativ
44	26	72.2	440	1	AM3D_ORYSA	P27932	oryza sativ
45	26	72.2	446	1	FD3C_ARATH	P46310	arabidopsis
46	26	72.2	454	1	ADPH_MYCTU	053320	mycobacteri
47	26	72.2	456	1	HMF1_DROXY	P48590	dtrosophila
48	26	72.2	457	1	KG90_HAEIN	P45320	haemophilus
49	26	72.2	470	1	ACHP_CHICK	P26153	gallus gall
50	26	72.2	475	1	LIPPL_HUMAN	P06858	homo sapien

ALIGNMENTS

RESULT	ID	CLD2_HUMAN	STANDARD:	PRT:	230 AA.
AC	P57739;				
DT	20-AUG-2001	(Rel. 40, Created)			
DT	20-AUG-2001	(Rel. 40, Last sequence update)			
DT	20-AUG-2001	(Rel. 40, Last annotation update)			
DE	CLAUDIN-2.				
GN	CLDN2.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
NCBI_TaxID=9606;					
RP	SEQUENCE FROM N.A.				
RC	Tissue=Epithelium;				
RA	Reinecker H.-C., Sakaguchi T., Golden H.M.;				
RL	Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.				
RM	[2]				
RP	SEQUENCE FROM N.A.				
RA	Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.O., Zhang P.P.,				
RA	Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.F., Li H.N., Yu Y.,				
RA	Yu J., Han L.H.;				
RT	"Novel human cDNA clone with function of inhibiting cancer cell				
RT	growth.";				
RL	Submitted (AUG-1999) to the EMBL/Genbank/DBJ databases.				
CC	- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.				
CC	- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.				
CC	- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isp-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
CC	EMBL: AF250558; AAF98151.1; -				
DR	EMBL: AF17340; AAG17984.1; -				
DR	InterPro: IPR001832; Claudin.				
DR	InterPro: IPR000729; PMP22.Claudin.				
DR	Pfam: PF00822; PMP22.Claudin; 1.				
DR	PROSITE: PS01346; CLAUDIN; 1.				
KM	Tight junction; Transmembrane.				
FT	TRANSMEM 8				
FT	TRANSMEM 82				
FT	TRANSMEM 117				
FT	TRANSMEM 137				
FT	TRANSMEM 163				
FT	TRANSMEM 183				
FT	TRANSMEM 230				
FT	SEQUENCE 230 AA; 24548 MW; 52CA642DA462B70D CRC64;				

```

Query Match      88.9%; Score 32; DB 1; Length 230;
Best Local Similarity 62.5%; Pred. NO. 3.5;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 wkxxyxg 8
      11 11 11
      30 WKTSSYVG 37

Db

RESULT 2
LHA2_ECTHL
ID      LHA2_ECTHL      STANDARD;      PRT;      65 AA.
AC      P80103;
DT      01-FEB-1994 (Rel. 28, Created)
DT      01-FEB-1994 (Rel. 28, Last sequence update)
DT      01-FEB-1994 (Rel. 28, Last annotation update)
DE      LIGHT-HARVESTING PROTEIN B800/830/1020, ALPHA-2 CHAIN (EHS-ALPHA-2)
OS      Ectochlorodospira halochloris.
OC      Bacteria; Proteobacteria; gamma subdivision; Ectochlorodospiraceae;
OC      Halorhodospira.
OC      NCBI_TaxID=1052;
OX      [1]
RN      SEQUENCE
RP      STRAIN=DSM 1059;
RX      MEDLINE=92249336; Pubmed=1577009;
RA      Wagner-Huber R., Brunisholz R.A., Bissig I., Frank G., Suter F.,
      Zuber H.;
RT      "The primary structure of the antenna polypeptides of
      Ectochlorodospira halochloris and Ectochlorodospira halophila. Four
      core-type antenna polypeptides in E. halochloris and E. halophila.";
RL      Eur. J. Biochem. 205:917-925(1992).
CC      -1- FUNCTION: ANTENNA COMPLEXES ARE LIGHT-HARVESTING SYSTEMS, WHICH
      TRANSFER THE EXCITATION ENERGY TO THE REACTION CENTERS.
      -1- SUBUNIT: THE CORE COMPLEX IS FORMED BY DIFFERENT ALPHA AND BETA
      CHAINS, BINDING BACTERIOCHLOROPHYLL MOLECULES, AND ARRANGED
      MOST PROBABLY IN TETRAMERIC STRUCTURES DISPOSED AROUND THE
      REACTION CENTER. THE NON-PIGMENTED GAMMA CHAINS MAY CONSTITUTE
      ADDITIONAL COMPONENTS.
CC      -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. INNER MEMBRANE.
CC      PIR: S23164; S23164.
DR      InterPro: IPR002361; Antenna_comp_alpha.
DR      InterPro: IPR000066; LHC.
DR      Pfam: PF00556; LHC; 1.
DR      PROSITE: PS00968; ANTENNA_COMP_ALPHA; 1.
KW      Antenna complex; Light-harvesting polypeptide; Transmembrane;
KW      Magnesium; Bacteriochlorophyll; Inner membrane;
FT      DOMAIN 1 13 CYTOPLASMIC (POTENTIAL).
FT      TRANSMEM 14 34 PERIPLASMIC (POTENTIAL).
FT      DOMAIN 35 65 AXIAL LIGAND TO THE BACTERIOCHLOROPHYLL.
FT      METAL 29 29 MAGNESIUM (POTENTIAL).
SO      SEQUENCE 65 AA; 7688 MW; 55A4C306748E3D9A CRC64;

Query Match      86.1%; Score 31; DB 1; Length 65;
Best Local Similarity 62.5%; Pred. NO. 1.8;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 wkxxyxg 8
      11 11 11
      30 WKTSSYDG 65

Db

RESULT 3
CLD1_MOUSE
ID      CLD1_MOUSE      STANDARD;      PRT;      211 AA.
AC      O88551;
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      CLAUDIN-1.

```

```

GN      CLDN1.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=98311639; Pubmed=9647647;
RA      Furuse M., Fujita K., Hiltrag T., Fujimoto K., Tsukita S.;
RT      "Claudin-1 and -2: novel integral membrane proteins localizing at
      tight junctions with no sequence similarity to occludin.";
RL      J. Cell Biol. 141:1539-1550(1998).
CC      -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC      -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC      -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: AF072127; AAC27078.1; -.
DR      MGD: MG1:1276109; Clnd1.
DR      InterPro: IPR001832; Claudin.
DR      InterPro: IPR000729; PMP22_Claudin.
DR      Pfam: PF00822; PMP22_Claudin; 1.
DR      PRINTS: PR01077; CLAUDIN.
DR      PROSITE: PS01346; CLAUDIN; 1.
KW      Tight junction; Transmembrane.
FT      TRANSMEM 8 28 POTENTIAL.
FT      TRANSMEM 82 102 POTENTIAL.
FT      TRANSMEM 116 136 POTENTIAL.
FT      TRANSMEM 164 184 POTENTIAL.
SO      SEQUENCE 211 AA; 22881 MW; BEF896FA62DBB6F0 CRC64;

Query Match      86.1%; Score 31; DB 1; Length 211;
Best Local Similarity 62.5%; Pred. NO. 5.4;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 wkxxyxg 8
      11 11 11
      30 WKTSSYAG 37

Db

RESULT 4
CLD1_RAT
ID      CLD1_RAT      STANDARD;      PRT;      211 AA.
AC      P56745;
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      CLAUDIN-1.
GN      CLDN1.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      STRAIN=SPRAGUE-DAWLEY;
RA      Gregory M., Dufresne J., Cyr D.G.;
RT      Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC      -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC      -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC      -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its

```



```

CC -----
DR EMBL: M10376; AAA46346.1; -.
DR PIR: A04157; Q0CV2.
SQ SEQUENCE 159 AA; 17787 MW; 70C76E75A6891ACF CRC64;

Query Match
Best Local Similarity 83.3%; Score 30; DB 1; Length 159;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxsyxg 8
   |||||
DB 61 WKINSYFG 68

RESULT 8
VAT_CAMVE
ID VAT_CAMVE STANDARD; PRT: 159 AA.
AC 002966;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE APHID TRANSMISSION PROTEIN (PROTEIN 2).
GN IT.
OS Cauliflower mosaic virus (strain BRC) (CamV).
OC Viruses; Retroid viruses; Caulimovirus.
OX NCBI_TaxID=31556;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93154593; PubMed=8428667;
RA Chenault K.D., Melcher U.K.;
RT "The complete nucleotide sequence of cauliflower mosaic virus isolate
RT BRC.";
RL Gene 123:255-257(1993).
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN VIRUS TRANSMISSION.
CC -!- SIMILARITY: BELONGS TO THE CAULIMOVIRUSES ORF II FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M90542; AAA62372.1; -.
DR PIR: JN0494; JN0494.
SQ SEQUENCE 159 AA; 17884 MW; 9B2012F605E5CA03 CRC64;

Query Match
Best Local Similarity 83.3%; Score 30; DB 1; Length 159;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxsyxg 8
   |||||
DB 61 WKINSYFG 68

RESULT 9
VAT_CAMVN
ID VAT_CAMVN STANDARD; PRT: 159 AA.
AC 000965;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE APHID TRANSMISSION PROTEIN (PROTEIN 2).
GN IT.
OS Cauliflower mosaic virus (strain NY8153) (CamV).
OC Viruses; Retroid viruses; Caulimovirus.
OX NCBI_TaxID=31557;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RA Chenault K.D., Steffens D.L., Melcher U.K.;
RT "Nucleotide sequence of cauliflower mosaic virus isolate NY8153.";
RL Plant Physiol. 100:342-345(1992).
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN VIRUS TRANSMISSION.
CC -!- SIMILARITY: BELONGS TO THE CAULIMOVIRUSES ORF II FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M90541; AAA46355.1; -.
DR PIR: A04157; Q0CV2.
SQ SEQUENCE 159 AA; 17775 MW; 346028DF17F5C6AE CRC64;

Query Match
Best Local Similarity 83.3%; Score 30; DB 1; Length 159;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxsyxg 8
   |||||
DB 61 WKINSYFG 68

RESULT 10
VAT_CAMVS
ID VAT_CAMVS STANDARD; PRT: 159 AA.
AC P03548;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE APHID TRANSMISSION PROTEIN (PROTEIN 2).
GN II.
OS Cauliflower mosaic virus (strain Strasbourg) (CamV).
OC Viruses; Retroid viruses; Caulimovirus.
OX NCBI_TaxID=10648;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=81001865; PubMed=7407912;
RA Franck A., Guillely H., Jonard G., Richards K., Hirth L.;
RT "Nucleotide sequence of cauliflower mosaic virus DNA.";
RL Cell 21:285-294(1980).
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN VIRUS TRANSMISSION.
CC -!- SIMILARITY: BELONGS TO THE CAULIMOVIRUSES ORF II FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: V00141; CAA23457.1; -.
DR PIR: A04157; Q0CV2.
SQ SEQUENCE 159 AA; 17803 MW; 748A326F17F5C1A9 CRC64;

Query Match
Best Local Similarity 83.3%; Score 30; DB 1; Length 159;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxsyxg 8
   |||||
DB 61 WKINSYFG 68

RESULT 11
CLD2_MOUSE
ID CLD2_MOUSE STANDARD; PRT: 230 AA.

```

```

AC 088552: 2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-2.
GN CLDN2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10099;
RN [1]
RP SEQUENCE FROM N.A.
RF MEDLINE=98311639; PubMed=9647647;
RA Furuse M., Fujita K., Hiragi T., Fujimoto K., Tsukita S.;
RT "Claudin-1 and -2: novel integral membrane proteins localizing at
RT tight junctions with no sequence similarity to occludin.";
RL J. Cell Biol. 141:1539-1550(1998).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) SPANS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF072128; AAC27079.1; -.
DR MGD; MGI:1276110; Cldn2.
DR InterPro; IPR001832; Claudin.
DR InterPro; IPR000729; PMP22.Claudin.
DR Pfam; PF008823; PMP22.Claudin; 1.
DR PRINTS; PR01077; CLAUDIN.
DR PROSITE; PS01346; CLAUDIN; 1.
DR 1Tight junction; Transmembrane.
KW TRANSMEM 8
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 163 183 POTENTIAL.
SQ SEQUENCE 230 AA; 24483 MW; 38A7C07A1E0D5D2 CRC64;

Query Match 80.6%; Score 29; DB 1; Length 230;
Best Local Similarity 50.0%; Pred. No. 16;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxxxxxg 8
   1: 111
Db 30 WRTSSYVG 37

RESULT 12
ID EX5B_ECOLI STANDARD; PRT; 1180 AA.
AC P08394;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EXODEOXYRIBONUCLEASE V BETA CHAIN (EC 3.1.11.5) (EXODEOXYRIBONUCLEASE
DE V 135 KDA POLYPEPTIDE).
DE RECB OR RORA OR B2820.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RF MEDLINE=87065729; PubMed=3537960;
RA Finch P.W., Storey A., Chapman K.E., Brown K., Hickson I.D.,
RA Emerson P.T.;
RT "Complete nucleotide sequence of the Escherichia coli recB gene.";
```

RL Nucleic Acids Res. 14:8573-8582(1986).
 RP SEQUENCE FROM N.A.
 RC STRAIN-V1000.
 RA MEDLINE=20229837; PubMed=10766864;
 RA Arnold D.A., Kowalczykowski S.C.;
 RT "Facilitated loading of RecA protein is essential to recombination by
 RT RecBCD enzyme."
 RL J. Biol. Chem. 275:12261-12265(2000).
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655.
 RA Blattner F.R., Plunkett G. II, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12."
 RL Science 277:1453-1474(1997).
 RP SEQUENCE OF 1-11 FROM N.A.
 RA MEDLINE=87040734; PubMed=3534791;
 RA Finch P.W., Wilson R.E., Brown K., Hickson I.D., Emerson P.T.;
 RT "Complete nucleotide sequence of the Escherichia coli ptr gene
 RT encoding pro tease III."
 RL Nucleic Acids Res. 14:7695-7703(1986).
 RP SEQUENCE OF 1093-1180 FROM N.A.
 RA MEDLINE=87065730; PubMed=3537961.
 RA Finch P.W., Storey A., Brown K., Hickson I.D., Emerson P.T.;
 RT "Complete nucleotide sequence of recD, the structural gene for the
 RT alpha subunit of Exonuclease V of Escherichia coli."
 RL Nucleic Acids Res. 14:8583-8594(1986).
 RP FUNCTION: REQUIRED FOR EFFICIENT DNA REPAIR; IT CATALYZES THE
 RP UNWINDING OF DOUBLE-STRANDED DNA AND THE CLEAVAGE OF SINGLE-
 RP STRANDED DNA AND IT STIMULATES LOCAL GENETIC RECOMBINATION.
 RP ALL OF THESE ACTIVITIES REQUIRE CONCOMITANT HYDROLYSIS OF ATP.
 RP -1 CATALYTIC ACTIVITY: EXONUCLEOTIDIC CLEAVAGE (IN THE PRESENCE OF
 RP ATP) IN EITHER 5'-TO 3'-OR 3'-TO 5'-DIRECTION TO YIELD 5'-
 RP PHOSPHOOLIGONUCLEOTIDES.
 RP -1 SUBUNIT: CONSIST OF THREE SUBUNITS; REC8, REC2 AND REC3.
 RP -1 SIMILARITY: BELONGS TO THE UVRD SUBFAMILY OF HELICASES.
 RP -----
 RP This SWISS-PROT entry is copyright. It is produced through a collaboration
 RP between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 RP the European Bioinformatics Institute. There are no restrictions on lists
 RP use by non-profit institutions as long as its content is in no way
 RP modified and this statement is not removed. Usage by and for commercial
 RP entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 RP or send an email to license@isb-sib.ch).
 RP -----
 RP EMBL: X04581: CAA28250.1; -
 RP EMBL: AF179304: AAD56369.1; -
 RP EMBL: U29581: AAB40467.1; -
 RP EMBL: AE000365: AAC75859.1; -
 RP EMBL: X06227: CAA29577.1; -
 RP EMBL: X04582: CAA28252.1; -
 RP PIR: A25532: NCECX5.
 RP HSSP: P56255: 1PUR.
 RP Ecogene: EG10824: recB.
 RP InterPro: IPR000212: UvrD-helicase.
 RP Pfam: PF00580: UvrD-helicase; 1.
 RP Hydrolase: Nuclease; Exonuclease; Endonuclease; Helicase; ATP-binding;
 RP DNA repair: Complete proteome.
 RP NP_BIND 23 ATP.
 RP FT 30
 RP SEQUENCE 1180 AA; 133958 MW; F9AC331808E8F281 CRC64;

Db 901 WRWTSYSG 908

RESULT 13

CLD7_RAT STANDARD; PRT; 191 AA.
AC 092111;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-7 (FRAGMENT).
GN CLDN7.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Keen T.J., Inglehearn C.F.; EMBL/GenBank/DBJ databases.
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: A011811; CA09790.1; -
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin; 1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN; 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 62 82 POTENTIAL.
FT TRANSMEM 100 120 POTENTIAL.
FT TRANSMEM 141 161 POTENTIAL.
SQ SEQUENCE 191 AA; 20366 MW; 83B445908DFFFA1A CRC64;

Query Match

Best Local Similarity 77.8%; Score 28; DB 1; Length 191;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxyxg 8
1: 111
DB 10 WQMSYAG 17

RESULT 14

CLD1_HUMAN STANDARD; PRT; 211 AA.
AC 095832;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-APR-2001 (Rel. 40, Last annotation update)
DE CLAUDIN-1 (SENESCENCE-ASSOCIATED EPITHELIAL MEMBRANE PROTEIN).
GN CLDN1 OR CLDI OR SEMPL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99132301; PubMed=9931503;
RA SwissProt=K.L., Machi A., Planitzer S., Robertson R., Kubbies M.,

RA Hosler S.;
RT "SEMP1, a senescence-associated cDNA isolated from human mammary
RT epithelial cells, is a member of an epithelial membrane protein
RT superfamily.";
RL Gene 226:285-295(1999).
RN [2]

RP SEQUENCE FROM N.A.
RA Mitic L.M., Anderson J.M.;
RT "Human claudin-1 isolated from Caco-2 mRNA."
RL Submitted (Dec-1998) to the EMBL/GenBank/DBJ databases.
RN [3]

RP SEQUENCE FROM N.A.
RX MEDLINE=20290992; PubMed=10828592;
RA Halford S., Spencer P., Greenwood J., Winton H., Hunt D.M.,

RA Adamson P.;
RT "Assignment(1) of claudin-1 (CLDN1) to human chromosome 3q28-->q29
RT with somatic cell hybrids."
RL Cytogenet. Cell Genet. 88:217-217(2000).

CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: AF101051; AAD16433.1; -
DR EMBL: AF115546; AAD22962.1; -
DR EMBL: AF134160; AAF61393.1; -
DR MIM: 603718; -

DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin; 1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN; 1.

KW Tight junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 116 136 POTENTIAL.
FT TRANSMEM 164 184 POTENTIAL.

FT TRANSMEM 164 184 POTENTIAL.
FT CONFLICT 62 62 I -> V (IN REF. 2).
FT CONFLICT 135 135 V -> A (IN REF. 2).
SQ SEQUENCE 211 AA; 22744 MW; 07269000E6C214F0 CRC64;

Query Match

Best Local Similarity 77.8%; Score 28; DB 1; Length 211;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxyxg 8
1: 111
DB 30 WRWTSYAG 37

RESULT 15

CLD7_HUMAN STANDARD; PRT; 211 AA.
AC 095471;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-7.
GN CLDN7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

RC Tissue-Colon adenocarcinoma.
RA Keen T.J.:
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION ('TJ') STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ011497; CA09626.1; -.
DR InterPro: IPR001832; Claudln.
DR Pfam: PF00822; PMP22_Claudin.1.
DR PRINTS: PRO1077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN.1.
KW Tight junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 118 138 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
SQ SEQUENCE 211 AA; 22390 MW; 7F3CC1B963D912E1 CRC64;

Query Match 77.8%; Score 28; DB 1; Length 211;
Best Local Similarity 50.0%; Pred. No. 25;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxsyxyg 8
I: I I I
Db 30 WQMSYAG 37

RESULT 16
CLD7_MOUSE STANDARD; PRT; 211 AA.
ID CLD7_MOUSE
AC 092261;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CLAUDIN-7.
GN CLDN7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RS SEQUENCE FROM N.A.
RA MEDLINE=99110921; PubMed=9892664;
RA Morita K., Furuse M., Fujimoto K., Tsukita S.;
RT "Claudin multigene family encoding four-transmembrane domain protein
RT components of tight junction strands";
RL Proc. Natl. Acad. Sci. U.S.A. 96:511-516(1999).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION ('TJ') STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF087825; AAD09760.1; -.
DR MGD: MGI:1859285; Clnd7.
DR InterPro: IPR001832; Claudln.

DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin.1.
DR PRINTS: PRO1077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN.1.
KW Tight junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 120 140 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
SQ SEQUENCE 211 AA; 22359 MW; 4FE87E3A57AC9F29 CRC64;

Query Match 77.8%; Score 28; DB 1; Length 211;
Best Local Similarity 50.0%; Pred. No. 25;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxsyxyg 8
I: I I I
Db 30 WQMSYAG 37

RESULT 17
CLDX_BRARE STANDARD; PRT; 215 AA.
ID CLDX_BRARE
AC 09YH92;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CLAUDIN-LIKE PROTEIN ZFA422 (CLAUSIN 7).
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RS SEQUENCE FROM N.A.
RA Keen T.J., Inglehearn C.F.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RS SEQUENCE FROM N.A.
RA MEDLINE=20525589; PubMed=11071763;
RA Chin A.J., Tsang M., Weinberg E.S.;
RT "Heart and gut chiralities are controlled independently from initial
RT heart position in the developing zebrafish.";
RL Dev. Biol. 227:403-421(2000).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION ('TJ') STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ011788; CA09776.1; -.
DR EMBL: AF260240; AAG24512.1; -.
DR InterPro: IPR001832; Claudln.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin.1.
DR PRINTS: PRO1077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN.1.
KW Tight junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 118 138 POTENTIAL.
FT TRANSMEM 163 183 POTENTIAL.
SQ SEQUENCE 215 AA; 22865 MW; BC04870B75B8CB9D CRC64;

Query Match 77.8%; Score 28; DB 1; Length 215;

Best Local Similarity 50.0%; Pred. No. 25;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxyxg 8
111111
Db 30 WKMSAYG 37

RESULT 18

APHA_MYCRA

STANDARD;

PRT; 341 AA.

AC 048935;

RC MEDLINE-21128732;

RT Nature 409.1007-1011(2001).

RL 15-JUL-1998 (Rel. 36, Created)

RT 15-JUL-1998 (Rel. 36, Last sequence update)

DE D-3-PHOSPHOGLYCERATE DEHYDROGENASE (EC 1.1.1.95) (PGDH).

GN SERA OR ML1692 OR MLCB637.25.

OS Mycobacterium leprae.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxID=1769;

ON [1]

RP SEQUENCE FROM N.A.

RC STRAIN-TN:

RT MEDLINE-21128732; PubMed-11234002;

RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,

RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,

RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,

RA Davies R.M., Devlin K., Duthoy S., Fellwell T., Fraser A., Hamlin N.,

RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,

RA Murphy L., Oliver K., Quail M.A., Rajandream M.-A., Rutherford K.M.,

RA Rutter S., Seeger K., Simon S., Simmonds M., Skellon J., Squares R.,

RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,

RA Bartell B.G.;

RT "Massive gene decay in the leprosy bacillus";

RL Nature 409.1007-1011(2001).

CC -1- CATALYTIC ACTIVITY: 3-PHOSPHOGLYCERATE + NAD(+) =

CC -1- 3-PHOSPHOHYDROXYPYRUVATE + NADH.

CC -1- PATHWAY: FIRST COMMITTED STEP IN THE "PHOSPHORYLATED" PATHWAY

CC OF L-SERINE BIOSYNTHESIS.

CC -1- SIMILARITY: BELONGS TO THE D-ISOMER SPECIFIC 2-HYDROXYACID

CC DEHYDROGENASES FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: Z99263; CAB16440.1; -

CC EMBL: AL583923; CAC30645.1; -

CC HSSP: P01542; ILCN.

CC Leproma; ML1692; -

CC DR InterPro: IPR002912; ACT.

CC DR InterPro: IPR002162; D_2_hydroxyacid_DH.

CC DR InterPro: IPR002055; NAD_Binding.

CC DR Pfam: PF00389; 2-Hacid_DH; 1.

CC DR Pfam: PF01842; ACT; 1.

CC DR PROSITE: PS00065; D_2-HYDROXYACID_DH_1; 1.

CC DR PROSITE: PS00670; D_2-HYDROXYACID_DH_2; 1.

CC DR PROSITE: PS00671; D_2-HYDROXYACID_DH_3; 1.

CC KW Serine biosynthesis: Oxidoreductase; NAD; Complete proteome.

CC FT ACT_SITE 232 232 SUBSTRATE-BINDING (BY SIMILARITY).

CC FT ACT_SITE 261 261 BY SIMILARITY.

CC FT ACT_SITE 279 279 BY SIMILARITY.

CC SQ SEQUENCE 528 AA; 54469 MW; 1A6DC609FB71222 CRC64;

CC -----

CC Query Match

CC Best Local Similarity 77.8%; Score 28; DB 1; Length 528;

CC Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

CC OY 1 wkxxyxg 8

CC 111111

CC Db 129 WKRSFSG 136

CC RESULT 20

CC SERA_MYCTU

CC ID SERA_MYCTU

CC AC 053243;

CC DT 15-JUL-1999 (Rel. 38, Created)

CC STANDARD;

CC PRT; 528 AA.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: Z99263; CAB16440.1; -

CC EMBL: AL583923; CAC30645.1; -

CC HSSP: P01542; ILCN.

CC Leproma; ML1692; -

CC DR InterPro: IPR002912; ACT.

CC DR InterPro: IPR002162; D_2_hydroxyacid_DH.

CC DR InterPro: IPR002055; NAD_Binding.

CC DR Pfam: PF00389; 2-Hacid_DH; 1.

CC DR Pfam: PF01842; ACT; 1.

CC DR PROSITE: PS00065; D_2-HYDROXYACID_DH_1; 1.

CC DR PROSITE: PS00670; D_2-HYDROXYACID_DH_2; 1.

CC DR PROSITE: PS00671; D_2-HYDROXYACID_DH_3; 1.

CC KW Serine biosynthesis: Oxidoreductase; NAD; Complete proteome.

CC FT ACT_SITE 232 232 SUBSTRATE-BINDING (BY SIMILARITY).

CC FT ACT_SITE 261 261 BY SIMILARITY.

CC FT ACT_SITE 279 279 BY SIMILARITY.

CC SQ SEQUENCE 528 AA; 54469 MW; 1A6DC609FB71222 CRC64;

CC -----

CC Query Match

CC Best Local Similarity 77.8%; Score 28; DB 1; Length 528;

CC Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

CC OY 1 wkxxyxg 8

CC 111111

CC Db 129 WKRSFSG 136

CC RESULT 20

CC SERA_MYCTU

CC ID SERA_MYCTU

CC AC 053243;

CC DT 15-JUL-1999 (Rel. 38, Created)

CC STANDARD;

CC PRT; 528 AA.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: Z99263; CAB16440.1; -

CC EMBL: AL583923; CAC30645.1; -

CC HSSP: P01542; ILCN.

CC Leproma; ML1692; -

CC DR InterPro: IPR002912; ACT.

CC DR InterPro: IPR002162; D_2_hydroxyacid_DH.

CC DR InterPro: IPR002055; NAD_Binding.

CC DR Pfam: PF00389; 2-Hacid_DH; 1.

CC DR Pfam: PF01842; ACT; 1.

CC DR PROSITE: PS00065; D_2-HYDROXYACID_DH_1; 1.

CC DR PROSITE: PS00670; D_2-HYDROXYACID_DH_2; 1.

CC DR PROSITE: PS00671; D_2-HYDROXYACID_DH_3; 1.

CC KW Serine biosynthesis: Oxidoreductase; NAD; Complete proteome.

CC FT ACT_SITE 232 232 SUBSTRATE-BINDING (BY SIMILARITY).

CC FT ACT_SITE 261 261 BY SIMILARITY.

CC FT ACT_SITE 279 279 BY SIMILARITY.

CC SQ SEQUENCE 528 AA; 54469 MW; 1A6DC609FB71222 CRC64;

CC -----

CC Query Match

CC Best Local Similarity 77.8%; Score 28; DB 1; Length 528;

CC Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

CC OY 1 wkxxyxg 8

CC 111111

CC Db 129 WKRSFSG 136

CC RESULT 20

CC SERA_MYCTU

CC ID SERA_MYCTU

CC AC 053243;

CC DT 15-JUL-1999 (Rel. 38, Created)

CC STANDARD;

CC PRT; 528 AA.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: Z99263; CAB16440.1; -

CC EMBL: AL583923; CAC30645.1; -

CC HSSP: P01542; ILCN.

CC Leproma; ML1692; -

CC DR InterPro: IPR002912; ACT.

CC DR InterPro: IPR002162; D_2_hydroxyacid_DH.

CC DR InterPro: IPR002055; NAD_Binding.

CC DR Pfam: PF00389; 2-Hacid_DH; 1.

CC DR Pfam: PF01842; ACT; 1.

CC DR PROSITE: PS00065; D_2-HYDROXYACID_DH_1; 1.

CC DR PROSITE: PS00670; D_2-HYDROXYACID_DH_2; 1.

CC DR PROSITE: PS00671; D_2-HYDROXYACID_DH_3; 1.

CC KW Serine biosynthesis: Oxidoreductase; NAD; Complete proteome.

CC FT ACT_SITE 232 232 SUBSTRATE-BINDING (BY SIMILARITY).

CC FT ACT_SITE 261 261 BY SIMILARITY.

CC FT ACT_SITE 279 279 BY SIMILARITY.

CC SQ SEQUENCE 528 AA; 54469 MW; 1A6DC609FB71222 CRC64;

CC -----

CC Query Match

CC Best Local Similarity 77.8%; Score 28; DB 1; Length 528;

CC Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

CC OY 1 wkxxyxg 8

CC 111111

CC Db 129 WKRSFSG 136

CC RESULT 20

CC SERA_MYCTU

CC ID SERA_MYCTU

CC AC 053243;

CC DT 15-JUL-1999 (Rel. 38, Created)

CC STANDARD;

CC PRT; 528 AA.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: Z99263; CAB16440.1; -

CC EMBL: AL583923; CAC30645.1; -

CC HSSP: P01542; ILCN.

CC Leproma; ML1692; -

CC DR InterPro: IPR002912; ACT.

CC DR InterPro: IPR002162; D_2_hydroxyacid_DH.

CC DR InterPro: IPR002055; NAD_Binding.

CC DR Pfam: PF00389; 2-Hacid_DH; 1.

CC DR Pfam: PF01842; ACT; 1.

CC DR PROSITE: PS00065; D_2-HYDROXYACID_DH_1; 1.

CC DR PROSITE: PS00670; D_2-HYDROXYACID_DH_2; 1.

CC DR PROSITE: PS00671; D_2-HYDROXYACID_DH_3; 1.

CC KW Serine biosynthesis: Oxidoreductase; NAD; Complete proteome.

DT 15-JUL-1999 (Rel. 38, last sequence update)
DT 20-AUG-2001 (Rel. 40, last annotation update)
DE D-3-PHOSPHOGLYCERATE DEHYDROGENASE (EC 1.1.1.95) (PGDH).
GN SERA OR RV2996C OR MT3074 OR MTV012.10.
OS Mycobacterium tuberculosis.
OC Bacteria: Firmicutes: Actinobacteria: Actinobacteridae:
OC Actinomycetales: Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID:1773;
NN [1]
NN SEQUENCE FROM N.A.

RP STRAIN-H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garner T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagsels K., Kiroh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RN SEQUENCE FROM N.A.

RP STRAIN-CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolony J.F., Nelson W.C., Umayan L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;

RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RT Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.

CC -1- CATALYTIC ACTIVITY: 3-PHOSPHOGLYCERATE + NAD(+) =

CC -1- 3-PHOSPHOHYDROXYPYRUVATE + NADH.

CC -1- PATHWAY: FIRST COMMITTED STEP IN THE "PHOSPHORYLATED" PATHWAY
CC OF L-SERINE BIOSYNTHESIS.

CC -1- SIMILARITY: BELONGS TO THE D-ISOMER SPECIFIC 2-HYDROXYACID
CC DEHYDROGENASES FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----

DR EMBL: AL021287; CAI16081.1; -
DR EMBL: AE007127; AKA47403.1; -
DR TIGR: M13074; -

DR TuberculList; RV2996c; -
DR InterPro: IPR002912; ACT.

DR InterPro: IPR002162; D_2_hydroxyacid_DH.
DR InterPro: IPR000205; NAD_binding.

DR Pfam: PF00389; 2-Hacid_DH; 1.
DR Pfam: PF01842; ACT; 1.

DR PROSITE: PS00065; D_2_HYDROXYACID_DH_1; 1.
DR PROSITE: PS00670; D_2_HYDROXYACID_DH_2; 1.

DR PROSITE: PS00671; D_2_HYDROXYACID_DH_3; 1.
KW Serine biosynthesis; Oxidoreductase; NAD; Complete proteome.

FT ACT_SITE 232 232 SUBSTRATE-BINDING (BY SIMILARITY).
FT ACT_SITE 261 261 BY SIMILARITY.

FT ACT_SITE 279 279 BY SIMILARITY.
FT ACT_SITE 279 279 BY SIMILARITY.

SO SEQUENCE 528 AA; 54554 MW; 3B5696A8D82A901 CRC64;

DB 11 1:1
129 WKRSSFSG 136

Search completed: January 14, 2002, 07:40:37
Job time: 503 sec

Query Match 77.8%; Score 28; DB 1; Length 528;
Best Local Similarity 50.0%; Pred. No. 59;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxyg 8

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:23:49 ; Search time 112.89 Seconds
(without alignments)
10.366 Million cell updates/sec

Title: 09-185908-1a
Perfect score: 36
Sequence: 1 wxxxxsyxg 8

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 50 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	88.9	193	11 Q9ET38	Q9et38 mus musculu
2	32	88.9	1083	2 Q9RPH6	Q9rph6 mycobacteri
3	31	86.1	432	5 Q9NFP2	Q9nfp2 plasmodium
4	30	83.3	64	12 Q83157	Q83157 cauliflower
5	30	83.3	99	12 Q83152	Q83152 cauliflower
6	30	83.3	159	12 Q83179	Q83179 cauliflower
7	30	83.3	159	12 Q66159	Q66159 cauliflower
8	30	83.3	159	12 Q83166	Q83166 cauliflower
9	30	83.3	159	12 Q9W133	Q9w133 cauliflower
10	30	83.3	246	5 Q44142	Q44142 caenorhabdi
11	29	80.6	65	2 Q9RD02	Q9rd02 streptomyce
12	29	80.6	65	2 Q9RD18	Q9rd18 streptomyce
13	29	80.6	65	2 Q9RD17	Q9rd17 streptomyce
14	29	80.6	94	12 Q91294	Q91294 human immun
15	29	80.6	94	12 Q91296	Q91296 human immun
16	29	80.6	94	12 Q91297	Q91297 human immun
17	29	80.6	96	12 Q91295	Q91295 human immun
18	29	80.6	100	12 Q9YKCL	Q9ymcl human immun
19	29	80.6	102	12 Q91298	Q91298 human immun

20	29	80.6	102	12 Q91299	Q91299 human immun
21	29	80.6	125	5 Q9NKN6	Q9nkn6 leishmania
22	29	80.6	134	12 Q9YLG3	Q9ylg3 human immun
23	29	80.6	149	12 Q9QLN6	Q9qln6 human immun
24	29	80.6	149	12 Q9QLN5	Q9qln5 human immun
25	29	80.6	174	5 Q77257	Q77257 dictyostell
26	29	80.6	635	5 Q9NKK0	Q9nkk0 leishmania
27	29	80.6	1208	2 Q9KPP6	Q9kpp6 vibrio chol
28	28	77.8	197	11 Q9JHG0	Q9jhg0 mus musculu
29	28	77.8	211	4 Q9BVN0	Q9bvn0 homo sapien
30	28	77.8	243	2 Q9ADD0	Q9add0 streptomyce
31	28	77.8	278	1 Q58458	Q58458 pyrococcus
32	28	77.8	308	12 Q9IR22	Q9ir22 dianthoviru
33	28	77.8	424	2 Q9RS14	Q9rs14 delnoccocus
34	28	77.8	429	2 Q9RDP7	Q9rdp7 streptomyce
35	28	77.8	475	10 Q9ZV90	Q9zv90 arabidopsis
36	28	77.8	475	10 Q9LP58	Q9lp58 arabidopsis
37	28	77.8	475	10 Q9ILR4	Q9ilr4 triticum ae
38	28	77.8	475	10 Q80774	Q80774 arabidopsis
39	28	77.8	488	1 Q28946	Q28946 archaeoglob
40	28	77.8	490	4 Q9NVC5	Q9nvc5 homo sapien
41	28	77.8	490	4 Q9H0R1	Q9h0r1 homo sapien
42	28	77.8	529	2 Q9Z564	Q9z564 streptomyce
43	28	77.8	1038	2 Q45780	Q45780 bacteroides
44	28	77.8	4152	2 Q9ZHL3	Q9zhl3 haemophilus
45	28	77.8	4919	2 Q9ZHL0	Q9zhl0 haemophilus
46	27	75.0	109	2 Q9PB67	Q9pb67 xylella fas
47	27	75.0	117	12 P89081	P89081 human adeno
48	27	75.0	121	4 Q9NUB9	Q9nub9 homo sapien
49	27	75.0	133	10 Q9C9C0	Q9c9c0 arabidopsis
50	27	75.0	149	4 Q9NUB8	Q9nub8 homo sapien

ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	193 AA.
Q9ET38	Q9ET38			
AC	Q9ET38:			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	CLAUDIN-19 (FRAGMENT) .			
OS	Mus musculus (Mouse) .			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ICR;			
RA	Kiuchi Y., Morita K., Furuse M., Tsukita S.;			
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF249889; AAF68323.1; -.			
DR	InterPro: IPR001832; Claudin.			
DR	InterPro: IPR000729; PMP22_Claudin.			
DR	Pfam: PF00822; PMP22_Claudin; 1.			
DR	PRINTS: PR01077; CLAUDIN.			
DR	PROSITE: PS01346; CLAUDIN; UNKNOWN_1.			
FT	NON_TER	1		
FT	NON_TER	193		
SQ	SEQUENCE	193 AA;	20299 MW;	2F2D82DB5FCF0D7F CRC64;
Query Match				
Best Local Similarity 88.9%; Score 32; DB 11; Length 193;				
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;				
Oy	1 wxxxxsyxg 8			
Db	20 WKOSSYAG 27			

```
RESULT 2
Q9RPH6 PRELIMINARY; PRT; 1083 AA.
AC Q9RPH6:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE RECB.
GN RECB.
OS Mycobacterium smegmatis.
OC Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1772;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC2155 (NB2);
RX MEDLINE=99412429; PubMed=10481025;
RA Griffin T.J. IV, Parsons L., Leschziner A.E., Devost J.,
RA Dethyshire K.M., Grindley N.D.F.;
RT "In vitro transposition of Tn552: a tool for DNA sequencing and
RT mutagenesis.";
RL Nucleic Acids Res. 27:3859-3865(1999).
DR EMBL: AF157643; AAD46808.1; -.
DR HSSP: P56255; 2PJR.
DR InterPro: IPR001993; Mitoch_carrier.
DR InterPro: IPR00212; UvrD_helicase.
DR Pfam: PF00580; UvrD_helicase; 1.
DR PROSITE: PS00215; MITOCH_CARRIER; UNKNOWN 1.
SQ SEQUENCE 1083 AA; 117442 MW; A531F5FBA85EB748 CRC64;
```

```
Query Match 88.9%; Score 32; DB 2; Length 1083;
Best Local Similarity 62.5%; Pred. No. 68;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 wxxsyxyg 8
|| || |
Db 766 WKRTSYSG 773
```

```
RESULT 3
Q9NEP2 PRELIMINARY; PRT; 432 AA.
AC Q9NEP2:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NIMA-RELATED PROTEIN KINASE (FRAGMENT).
GN NEK-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Dorin D., Le Roch K., Scallafandro P., Alano P., Parzy D., Meijer L.,
RA Doeris C., a novel NIMA-related protein kinase from the human malaria
RA "Pfkne-1, a novel NIMA-related protein kinase from the human malaria
RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RL -1- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC EMBL: AJ276023; CAB76949.1; -.
DR InterPro: IPR000719; Euk_kinase.
DR InterPro: IPR002290; Ser_thr_kin_actsite.
DR Pfam: PF00069; Kinase; 1.
DR SMART: SM00220; S_TKC; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
KM NON_TER 432 432
SQ SEQUENCE 432 AA; 49698 MW; C5FFE06A7C929E5D CRC64;
```

Query Match 86.1%; Score 31; DB 5; Length 432;

```
Best Local Similarity 62.5%; Pred. No. 43;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 wxxsyxyg 8
|| || |
Db 43 WKATSYRG 50
```

```
RESULT 4
Q83157 PRELIMINARY; PRT; 64 AA.
ID Q83157
AC Q83157:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE APHID ACQUISITION FACTOR 5' END (FRAGMENT).
OS Cauliflower mosaic virus.
OC Viruses; Retroid viruses; Caulimovirus.
OX NCBI_TaxID=10641;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90320145; PubMed=2371775;
RA Vaden V.R., Melcher U.;
RT "Recombination sites in cauliflower mosaic virus DNAs: implications
RT for mechanisms of recombination.";
RL Virology 177:717-726(1990).
DR EMBL: M32809; AAA46340.1; -.
FT NON_TER 1 1
FT NON_TER 64 64
SQ SEQUENCE 64 AA; 7109 MW; A20E7E25C93245FA CRC64;
```

```
Query Match 83.3%; Score 30; DB 12; Length 64;
Best Local Similarity 62.3%; Pred. No. 9.5;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 wxxsyxyg 8
|| || |
Db 7 WKINSYRG 14
```

```
RESULT 5
Q83162 PRELIMINARY; PRT; 99 AA.
ID Q83162
AC Q83162:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-NOV-1996 (TREMBLrel. 01, Last annotation update)
DE HYPOTHETICAL PROTEIN (FRAGMENT).
OS Cauliflower mosaic virus.
OC Viruses; Retroid viruses; Caulimovirus.
OX NCBI_TaxID=10641;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=APHID NONTRANSMISSIBLE CAMPBELL ISOLATE;
RX MEDLINE=87122167; PubMed=3027976;
RA Woolston C.J., Czaplewski L.G., Markham P.G., Goad A.S., Hull R.,
RA Davies J.W.;
RT "Characterization of a virus-specific proteolytic activity processing
RT the gag precursor of the simian sarcoma-associated virus.";
RL Virology 160:246-251(1987).
DR EMBL: M17415; AAA66604.1; -.
DR Hypothetical protein.
KM NON_TER 99 99
SQ SEQUENCE 99 AA; 11146 MW; CFF8E64E82533CD6 CRC64;
```

```
Query Match 83.3%; Score 30; DB 12; Length 99;
Best Local Similarity 62.5%; Pred. No. 15;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

OY 1 wxxsyxyg 8
|| || |

```
Db 6Y WKINSYFG 68

RESULT 6
ID 083179 PRELIMINARY: PRT: 159 AA.
AC 083179;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMBlrel. 01, Last annotation update)
DE ORF II PROTEIN.
OS Cauliflower mosaic virus.
OC Viruses; Retroid viruses; Caulimovirus.
NCBI_TaxID=10641;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC1841.
RA MEDLINE=85179981; PubMed:2965429;
RT "Molecular mechanisms regulating the synthesis of transferin
receptors and ferritin in human erythroleukemic cell lines.";
RL FBS left. 181:223-228(1985).
DR EMBL: M37582; AAA96696.1; -.
SQ SEQUENCE 159 AA; 17941 MW; FF7D27C67CB38D41 CRC64;

Query Match 83.3%; Score 30; DB 12; Length 159;
Best Local Similarity 62.5%; Pred. No. 25;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxyxg 8
DB 61 WKINSYFG 68

RESULT 7
ID 066159 PRELIMINARY: PRT: 159 AA.
AC 066159;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMBlrel. 08, Last annotation update)
DE ORF II.
OS Cauliflower mosaic virus.
OC Viruses; Retroid viruses; Caulimovirus.
NCBI_TaxID=10641;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=829;
RA MEDLINE=95237629; PubMed:7721109;
RT Plaque M., Mougnot J.L., Geldreich A., Guldasci T., Mesnard J.M.,
Lebourlier G., Yot P.;
RT "Sequence of a cauliflower mosaic virus strain infecting solanaceous
plants.";
RL Gene 155:305-306(1995).
DR EMBL: X79465; CAAS5971.1; -.
SQ SEQUENCE 159 AA; 17759 MW; 752131389A2DFA39 CRC64;

Query Match 83.3%; Score 30; DB 12; Length 159;
Best Local Similarity 62.5%; Pred. No. 25;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxyxg 8
DB 61 WKINSYFG 68

RESULT 8
ID 083166 PRELIMINARY: PRT: 159 AA.
AC 083166;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
```

```
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMBlrel. 08, Last annotation update)
DE APHID TRANSMISSION PROTEIN.
OS Cauliflower mosaic virus.
OC Viruses; Retroid viruses; Caulimovirus.
NCBI_TaxID=10641;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=9413488; PubMed:8310068;
RA Chenault K.D., Melcher U.;
RT "Cauliflower mosaic virus isolate CMV-1.";
RL Plant Physiol. 101:1395-1396(1993).
DR EMBL: M90543; AAA21733.1; -.
SQ SEQUENCE 159 AA; 17815 MW; 753AA2F7AEFFC109 CRC64;

Query Match 83.3%; Score 30; DB 12; Length 159;
Best Local Similarity 62.5%; Pred. No. 25;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxyxg 8
DB 61 WKINSYFG 68

RESULT 9
ID 09W133 PRELIMINARY: PRT: 159 AA.
AC 09W133;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE APHID TRANSMISSION HELPER FACTOR.
GN ORF2.
OS Cauliflower mosaic virus.
OC Viruses; Retroid viruses; Caulimovirus.
NCBI_TaxID=10641;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=XINJIANG;
RA Pang R., Wu X., Bu M., Tian Y., Cai F., Mang K.;
RT "Complete nucleotide sequence of cauliflower mosaic virus (Xinjiang
RT isolate) genomic DNA.";
RL Ping Tu Hsueh Pao 1:247-256(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=XINJIANG;
RA Pang R., Wu X., Bu M., Tian Y., Cai F., Mang K.;
RL Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF140604; AAD37338.1; -.
SQ SEQUENCE 159 AA; 17816 MW; 753AA2F7AEFFC1A9 CRC64;

Query Match 83.3%; Score 30; DB 12; Length 159;
Best Local Similarity 62.5%; Pred. No. 25;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxyxg 8
DB 61 WKINSYFG 68

RESULT 10
ID 04A142 PRELIMINARY: PRT: 246 AA.
AC 04A142;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE C44B12.3 PROTEIN.
GN C44B12.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
```

```

OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
EX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., Mumurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Spratt J., Wohlman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Tin-Mollam A.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF036692; AAB88326.1; -;
SQ SEQUENCE 246 AA; 28343 MW; 33FE5E33ACFD55D3 CRC64;

Query Match      83.3%; Score 30; DB 5; Length 246;
Best Local Similarity 62.5%; Pred. No. 40;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxssyxxg 8
   | | | |
Db 87 WKXSSYSG 94

RESULT 11
O9RD02 PRELIMINARY; PRT; 65 AA.
AC O9RD02;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE HYPOTHEMETICAL 6.9 KDA PROTEIN.
GN SC447.10.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
MDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmid and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."

```

```

RL Mol. Microbiol. 21:77-96(1996).
DR EMBL: AL133423; CAB62714.1; -.
KW Hypothetical protein.
SQ SEQUENCE 65 AA; 6944 MW; F283FA15A0650DCE CRC64;

Query Match      80.6%; Score 29; DB 2; Length 65;
Best Local Similarity 50.0%; Pred. No. 16;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxssyxxg 8
   | | | |
Db 12 WKXSSYSG 19

RESULT 12
O9RD18 PRELIMINARY; PRT; 65 AA.
AC O9RD18;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE HYPOTHEMETICAL 7.0 KDA PROTEIN.
GN SCC57A.09C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
MDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmid and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RT Mol. Microbiol. 21:77-96(1996).
DR EMBL: AL136519; CAB6277.1; -.
KW Hypothetical protein.
SQ SEQUENCE 65 AA; 6959 MW; 1F74C265B9572610 CRC64;

Query Match      80.6%; Score 29; DB 2; Length 65;
Best Local Similarity 50.0%; Pred. No. 16;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxssyxxg 8
   | | | |
Db 7 WKXSSYSG 14

RESULT 13
O9RD17 PRELIMINARY; PRT; 65 AA.
AC O9RD17;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE HYPOTHEMETICAL 7.1 KDA PROTEIN.
GN SCC57A.10C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

```

OX NCBI_TaxID:1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN A3(2);
RX MEDLINE:97000351; PubMed=8843436;
RA Redenbach M., Kleeser H.M., Denepite D., Eichner A., Cullum J.,
RT Kinashi H., Hopwood D.A.;
RA "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL: AL136519; CAB6278.1; -
KM Hypothetical protein.
SQ SEQUENCE 65 AA: 7127 MW: 4C9A4476C44B727A CRC64;

Query Match 80.6%; Score 29; DB 2; Length 65;
Best Local Similarity 50.0%; Pred. No. 16;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxyxg 8
DB 7 WRKSSYSG 14

RESULT 14
ID 091294 PRELIMINARY; PRT; 94 AA.
AC 091294;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID:11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT PA FROM FRANCE;
RX MEDLINE:98285741; PubMed=9621043;
RA Belc L., Si Mohamed A., Muller-Trutwin M.C., Gillaquin J., Gutmann L.,
RT Safar M., Barre-Sinoussi F., Kazatchkine M.D.;
RA "Genetically related human immunodeficiency virus type 1 in three
RT adults of a family with no identified risk factor for intrafamilial
RT transmission.";
RL J. Virol. 72:5831-5839(1998).
DR EMBL: U87181; AAC32941.1; -
DR InterPro: IPR000777; GP120.
DR Pfam: PF00516; GP120; 1.
KM Envelope protein.
FT NON_TER 1
FT NON_TER 94
SQ SEQUENCE 94 AA: 10572 MW: 08A531FC4D023A74 CRC64;

Query Match 80.6%; Score 29; DB 12; Length 94;
Best Local Similarity 50.0%; Pred. No. 24;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
OY 1 wxxxxxyxg 8
DB 20 WKSNTYTG 27

RESULT 15
ID 091296 PRELIMINARY; PRT; 94 AA.
AC 091296;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID:11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT PA FROM FRANCE;
RX MEDLINE:98285741; PubMed=9621043;
RA Belc L., Si Mohamed A., Muller-Trutwin M.C., Gillaquin J., Gutmann L.,
RT Safar M., Barre-Sinoussi F., Kazatchkine M.D.;
RA "Genetically related human immunodeficiency virus type 1 in three
RT adults of a family with no identified risk factor for intrafamilial
RT transmission.";
RL J. Virol. 72:5831-5839(1998).
DR EMBL: U87183; AAC32943.1; -
DR InterPro: IPR000777; GP120.
DR Pfam: PF00516; GP120; 1.
KM Envelope protein.
FT NON_TER 1
FT NON_TER 94
SQ SEQUENCE 94 AA: 10550 MW: B4BE1E3D8193A10 CRC64;

Query Match 80.6%; Score 29; DB 12; Length 94;
Best Local Similarity 50.0%; Pred. No. 24;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxyxg 8
DB 20 WKSNTYTG 27

RESULT 16
ID 091297 PRELIMINARY; PRT; 94 AA.
AC 091297;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID:11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT PA FROM FRANCE;
RX MEDLINE:98285741; PubMed=9621043;
RA Belc L., Si Mohamed A., Muller-Trutwin M.C., Gillaquin J., Gutmann L.,
RT Safar M., Barre-Sinoussi F., Kazatchkine M.D.;
RA "Genetically related human immunodeficiency virus type 1 in three
RT adults of a family with no identified risk factor for intrafamilial
RT transmission.";
RL J. Virol. 72:5831-5839(1998).
DR EMBL: U87184; AAC32944.1; -
DR InterPro: IPR000777; GP120.
DR Pfam: PF00516; GP120; 1.
KM Envelope protein.
FT NON_TER 1
FT NON_TER 94
SQ SEQUENCE 94 AA: 10707 MW: 68AAED72CA1D73B9 CRC64;

Query Match 80.6%; Score 29; DB 12; Length 94;

Best Local Similarity 50.0%; Pred. No. 24;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsxyg 8
11:11
Db 20 WKSNTYG 27

RESULT 17

091295 PRELIMINARY; PRT; 96 AA.
AC 091295; 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT PA FROM FRANCE;
RX MEDLINE=98285741; PubMed=9621043;
RA Belc L., Si Mohamed A., Muller-Trutwin M.C., Gilquin J., Gutmann L.,
Safar M., Barre-Sinoussi F., Kazatchkine M.D.;
RT "Genetically related human immunodeficiency virus type 1 in three
adults of a family with no identified risk factor for intrafamilial
transmission."
RT J. Virol. 72:5831-5839(1998).
EMBL; U87182; AAC32942.1; -.
DR InterPro: IPR000777; GP120.
DR Pfam: PF00516; GP120; 1.
KW Envelope protein.
FT NON_TER 1 1
FT NON_TER 96 96
SQ SEQUENCE 96 AA; 10793 MW; 3415BAF6189073AA CRC64;

Query Match 80.6%; Score 29; DB 12; Length 96;
Best Local Similarity 50.0%; Pred. No. 25;

Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsxyg 8
11:11
Db 22 WKSNTYG 29

RESULT 18

091295 PRELIMINARY; PRT; 100 AA.
AC 091295; 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT PA FROM FRANCE;
RX MEDLINE=98285741; PubMed=9621043;
RA Belc L., Si Mohamed A., Muller-Trutwin M.C., Gilquin J., Gutmann L.,
Safar M., Barre-Sinoussi F., Kazatchkine M.D.;
RT "Genetically related human immunodeficiency virus type 1 in three
adults of a family with no identified risk factor for intrafamilial
transmission."
RT J. Virol. 72:5831-5839(1998).
EMBL; U87186; AAC32946.1; -.
DR InterPro: IPR000777; GP120.
DR Pfam: PF00516; GP120; 1.

KW Envelope protein.
FT NON_TER 1 1
FT NON_TER 100 100
SQ SEQUENCE 100 AA; 11199 MW; 904454395BAD30AD CRC64;

Query Match 80.6%; Score 29; DB 12; Length 100;
Best Local Similarity 50.0%; Pred. No. 26;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsxyg 8
11:11
Db 26 WKSNTYG 33

RESULT 19

091298 PRELIMINARY; PRT; 102 AA.
AC 091298; 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT PA FROM FRANCE;
RX MEDLINE=98285741; PubMed=9621043;
RA Belc L., Si Mohamed A., Muller-Trutwin M.C., Gilquin J., Gutmann L.,
Safar M., Barre-Sinoussi F., Kazatchkine M.D.;
RT "Genetically related human immunodeficiency virus type 1 in three
adults of a family with no identified risk factor for intrafamilial
transmission."
RT J. Virol. 72:5831-5839(1998).
EMBL; U87187; AAC32947.1; -.
DR InterPro: IPR000777; GP120.
DR Pfam: PF00516; GP120; 1.
KW Envelope protein.
FT NON_TER 1 1
FT NON_TER 102 102
SQ SEQUENCE 102 AA; 11399 MW; 8BB041793D708106 CRC64;

Query Match 80.6%; Score 29; DB 12; Length 102;
Best Local Similarity 50.0%; Pred. No. 26;

Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsxyg 8
11:11
Db 28 WKSNTYG 35

RESULT 20

091299 PRELIMINARY; PRT; 102 AA.
AC 091299; 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT PA FROM FRANCE;
RX MEDLINE=98285741; PubMed=9621043;
RA Belc L., Si Mohamed A., Muller-Trutwin M.C., Gilquin J., Gutmann L.,
Safar M., Barre-Sinoussi F., Kazatchkine M.D.;

RT "Genetically related human immunodeficiency virus type 1 in three
RT adults of a family with no identified risk factor for intrafamilial
transmission." J. Virol. 72:5831-5839(1998).
RL EMBL: U87188; AAC32948.1; -
DR InterPro: IPR000777; GP120.
DR Pfam: PF00516; GP120; 1.
KW Envelope protein.
FT NON_TER 1 1
FT NON_TER 102 102
SQ SEQUENCE 102 AA: 11399 MW: 888041793D/08106 CRC64;

Query Match 80.6%; Score 29; DB 12; Length 102;
Best Local Similarity 50.0%; Pred. No. 26;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Oy 1 WKKXXYXG 8
11 : 11
Db 28 WKGSNYTG 35

Search completed: January 14, 2002, 07:39:39
Job time: 950 sec

1

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:22:29 ; Search time 103.51 Seconds
(without alignments)
5.725 Million cell updates/sec

Title: 09-185908-1A
Perfect score: 36
Sequence: 1 wkxxyxg 8

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database :
1: A_Geneseq_1101: *
2: /SID52/gcgdata/geneseq/geneseqp/AA1980.DAT: *
3: /SID52/gcgdata/geneseq/geneseqp/AA1981.DAT: *
4: /SID52/gcgdata/geneseq/geneseqp/AA1982.DAT: *
5: /SID52/gcgdata/geneseq/geneseqp/AA1983.DAT: *
6: /SID52/gcgdata/geneseq/geneseqp/AA1984.DAT: *
7: /SID52/gcgdata/geneseq/geneseqp/AA1985.DAT: *
8: /SID52/gcgdata/geneseq/geneseqp/AA1986.DAT: *
9: /SID52/gcgdata/geneseq/geneseqp/AA1987.DAT: *
10: /SID52/gcgdata/geneseq/geneseqp/AA1988.DAT: *
11: /SID52/gcgdata/geneseq/geneseqp/AA1989.DAT: *
12: /SID52/gcgdata/geneseq/geneseqp/AA1990.DAT: *
13: /SID52/gcgdata/geneseq/geneseqp/AA1991.DAT: *
14: /SID52/gcgdata/geneseq/geneseqp/AA1992.DAT: *
15: /SID52/gcgdata/geneseq/geneseqp/AA1993.DAT: *
16: /SID52/gcgdata/geneseq/geneseqp/AA1994.DAT: *
17: /SID52/gcgdata/geneseq/geneseqp/AA1995.DAT: *
18: /SID52/gcgdata/geneseq/geneseqp/AA1996.DAT: *
19: /SID52/gcgdata/geneseq/geneseqp/AA1997.DAT: *
20: /SID52/gcgdata/geneseq/geneseqp/AA1998.DAT: *
21: /SID52/gcgdata/geneseq/geneseqp/AA1999.DAT: *
22: /SID52/gcgdata/geneseq/geneseqp/AA2000.DAT: *
23: /SID52/gcgdata/geneseq/geneseqp/AA2001.DAT: *

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	32	88.9	55	20	AAV12227
2	32	88.9	56	22	AAE04286
3	32	88.9	83	22	AAE0407
4	32	88.9	113	22	AAE04281
5	32	88.9	114	20	AAV12226
6	32	88.9	126	22	AAE04282
7	32	88.9	126	22	AAE04283
8	32	88.9	140	22	AAE04228
9	32	88.9	155	21	AAE04052
10	32	88.9	230	20	AAV36134
11	32	88.9	230	20	AAV36181

12	32	88.9	230	21	AAV99378	Human PRO1356 (UNQ
13	32	88.9	230	21	AAV84609	A human membrane a
14	32	88.9	230	22	AAE04286	Human polypeptide
15	32	88.9	230	22	AAU12417	Human PRO1356 poly
16	32	88.9	230	22	AAE04207	Human gene 10 enco
17	32	88.9	230	22	AAE04207	Human PRO1356, Ho
18	32	88.9	230	22	AAE04207	Human membrane or
19	32	88.9	230	22	AAE04207	Protein of the Inv
20	32	88.9	260	22	AAE04207	Human polypeptide
21	31	86.1	8	21	AAE04207	Claudin-1 cell adh
22	31	86.1	9	21	AAE04207	Claudin-1 cell adh
23	31	86.1	9	21	AAE04207	Claudin-1 cell adh
24	31	86.1	10	21	AAE04207	Claudin-1 cell adh
25	31	86.1	10	21	AAE04207	Claudin-1 cell adh
26	31	86.1	10	21	AAE04207	Claudin-1 cell adh
27	31	86.1	10	21	AAE04207	Claudin-1 cell adh
28	31	86.1	10	21	AAE04207	Claudin-1 cell adh
29	31	86.1	10	21	AAE04207	Claudin-1 cell adh
30	31	86.1	211	21	AAV51675	Murine clodin 1 pr
31	31	83.3	84	20	AAV06346	EGTII-like cellula
32	29	80.6	8	21	AAE06521	Claudin-2 cell adh
33	29	80.6	8	21	AAE06521	Claudin-2 cell adh
34	29	80.6	10	21	AAE06530	Claudin-2 cell adh
35	29	80.6	10	21	AAE06539	Claudin-2 cell adh
36	29	80.6	10	21	AAE06548	Claudin-2 cell adh
37	29	80.6	10	21	AAE06557	Claudin-2 cell adh
38	29	80.6	10	21	AAE06566	Claudin-2 cell adh
39	29	80.6	230	21	AAV51675	Murine clodin 2 pr
40	28	77.8	8	21	AAE06426	Claudin-1 cell adh
41	28	77.8	8	21	AAE06512	Claudin-1 cell adh
42	28	77.8	8	21	AAE06823	Claudin-7 cell adh
43	28	77.8	8	21	AAE06876	Claudin-7 cell adh
44	28	77.8	8	21	AAE06917	Claudin-7 cell adh
45	28	77.8	9	21	AAE06427	Claudin-1 cell adh
46	28	77.8	10	21	AAE06485	Claudin-1 cell adh
47	28	77.8	10	21	AAE06491	Claudin-1 cell adh
48	28	77.8	10	21	AAE06497	Claudin-1 cell adh
49	28	77.8	10	21	AAE06503	Claudin-1 cell adh
50	28	77.8	10	21	AAE06509	Claudin-1 cell adh

ALIGNMENTS

RESULT 1	AAV12227	standard; Protein; 55 AA.
ID	AAV12227	
XX	AAV12227	
AC	AAV12227	
XX	AAV12227	
DT	18-JUN-1999	(first entry)
XX		
XX		
DE	Human 5' EST secreted protein SEQ ID NO: 540.	
XX		
KW	Human; secreted protein; EST; expressed sequence tag; diagnosis;	
KW	forensic; gene therapy; chromosome mapping; signal peptide;	
KW	upstream regulatory sequence; cytokine activity; cell proliferation;	
KW	differentiation; haematopoiesis regulation; tissue growth regulation;	
KW	reproductive hormone regulation; chemokine; chemokine; haemostatic;	
KW	thrombolytic; anti-inflammatory; tumour inhibition.	
XX		
OS	Homo sapiens.	
XX		
PN	W09906554-A2.	
XX		
PD	11-FEB-1999.	
XX		
PF	31-JUL-1998; 98WO-1B01238.	
XX		
PR	01-AUG-1997; 97US-0905134.	
XX		
PA	(GEST) GENSET.	

XX Duclert A, Dumas Milne Edwards J, Lacroix B;
PI WPI: 1999-153784/13.
XX N-PSDB; AAX41060.
DR
XX New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from kidney, fetal kidney, dystrophic
PT muscle, muscle and heart tissue
XX
PS Claim 34; Page 601; 622pp; English.
XX
CC AAX40826 to AAX41093 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins, and encode the proteins given in AAX01602 and
CC AAX11994 to AAX12260, respectively. The proteins given represent the
CC signal peptide and an N-terminal fragment of a secreted protein. The
CC nucleic acid sequences can be used for producing secreted human gene
CC products. They can also be used to develop products for diagnosis and
CC therapy. The proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used
CC for directing extracellular secretion of a polypeptide or the insertion
CC of a polypeptide into a membrane, or importing a polypeptide into
CC a cell.
XX
SQ Sequence 55 AA;
OY
DB 1 wkxxsyxg 8
30 wktssyvg 37
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Query Match 88.9%; Score 32; DB 20; Length 55;
Best Local Similarity 62.5%; Pred. No. 8.3;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
RESULT 2
ID AAE04286 standard; peptide; 56 AA.
XX
AC AAE04286;
DT 09-AUG-2001 (first entry)
XX
DE Human gene 10 encoded secreted protein fragment, SEQ ID NO:150.
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angio-genic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;
KW endocrine disorder; infection; wound healing; vulnery;
KW cell culture; chemotaxis; food additive;
KW binding partner identification.
XX
OS Homo sapiens.
XX
PN WO200136432-A2.
XX
PD 25-MAY-2001.
XX
PF 15-NOV-2000; 2000WO-US31162.
XX

PR 19-NOV-1999; 99US-0166415.
PR 30-JUN-2000; 2000US-0215136.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis GA, Baker KP, Young PE;
XX WPI: 2001-343793/36.
DR
XX
PT Isolated nucleic acid molecule encoding a human secreted protein 1s
PT used in preventing, treating or ameliorating a medical condition -
XX
XX Disclosure; Page 39; 509pp; English.
XX
CC AAN08488-AAD08529 represent cDNAs corresponding to 18 human secreted
CC protein genes, and AAE04199-AAE04239 represent the proteins they encode.
CC AAE04240-AAE04297 represent human secreted protein fragments or variants.
CC The secreted proteins and their genes are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 18 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angio-genic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin ageing due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein fragment referred to in the disclosure of the invention.
XX
SQ Sequence 56 AA;
OY
DB 1 wkxxsyxg 8
6 wktssyvg 13
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Query Match 88.9%; Score 32; DB 22; Length 56;
Best Local Similarity 62.5%; Pred. No. 8.4;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
RESULT 3
ID AAM40407 standard; Protein; 83 AA.
XX
AC AAM40407;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 3552.
XX
XX Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX

OY 1 wkxxyxg 8
11 11 1
DB 3 wkssyvg 10

RESULT 5

AAV12226
ID AAV12226 standard; Protein; 114 AA.

AC AAV12226;

DT 18-JUN-1999 (first entry)

DE Human 5' EST secreted protein SEQ ID NO: 539.

KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition.

OS Homo sapiens.

PN MO9906554-A2.

PD 11-FEB-1999.

PF 31-JUL-1998; 98MO-IB01238.

PR 01-AUG-1997; 97US-0905134.

PA (GEST) GENSET.

PI Duclert A, Dumas Milne Edwards J, Lacroix B;

DR WPI: 1999-153784/13.

DR N-PSDB: AAX41059.

PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from Kidney, fetal kidney, dystrophic
PT muscle, muscle and heart tissue

PS Claim 34; Page 600-601; 622pp; English.

CC AAX40826 to AAX41093 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins, and encode the proteins given in AAY01602 and
CC AAY11994 to AAY12260, respectively. The proteins given represent the
CC signal peptide and an N-terminal fragment of a secreted protein. The
CC nucleic acid sequences can be used for producing secreted human gene
CC products. They can also be used to develop products for diagnosis and
CC therapy. The proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used
CC for directing extracellular secretion of a polypeptide or the insertion
CC of a polypeptide into a membrane, or importing a polypeptide into
CC a cell.

XX Sequence 114 AA;

Query Match 88.9%; Score 32; DB 20; Length 114;
Best Local Similarity 62.5%; Pred. No. 17;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxyxg 8
11 11 1
DB 30 wkssyvg 37

RESULT 6

AAM25829
ID AAM25829 standard; Protein; 126 AA.

AC AAM25829;

DT 16-OCT-2001 (first entry)

DE Human protein sequence SEQ ID NO:1344.

KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW antiinflammatory; antihemetic; antiallergic; immunosuppressive;
KW antibacterial; endocrine; cardiant; central nervous system; virucide;
KW anti-HIV; fungicide; antimutagen; cardiovascular; antinaemic; anaemia;
KW antiaggregant; haemostatic; vulnary; antilucer; osteopathic; eczema;
KW dermatological; antiallergic; antidiressant; noctropic; antiparkinsonian; infection;
KW immunoprotective; gene therapy; antisense therapy; vaccine; inflammation;
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW neurological disorder.

OS Homo sapiens.

PN MO200153455-A2.

PD 26-JUL-2001.

PF 22-DEC-2000; 2000MO-US35017.

PR 23-DEC-1999; 99US-0471275.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

DR WPI: 2001-457603/49.

DR N-PSDB: AAH99770.

PT Isolated human polynucleotides encoding polypeptides, useful for the
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -

PS Claim 20; Page 278; 1217pp; English.

CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC AAM25963. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
CC antiallergic; immunosuppressive; antibacterial; endocrine; cardiant;
CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC cardiovascular; antinaemic; antiaggregant; haemostatic; vulnary;
CC antilucer; osteopathic; dermatological; antiallergic; antiasthmatic;
CC antidiabetic; cytostatic; neuroprotective; antidiressant; noctropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic

CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.
XX
SQ Sequence 126 AA:

Query Match 88.9%; Score 32; DB 22; Length 126;
Best Local Similarity 62.5%; Pred. No. 19;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsyxg 8
|| || |
DB 73 wkqsyag 80

RESULT 7
AAM42193
ID AAM42193 standard; Protein: 126 AA.

AC AAM42193;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 7124.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW leukoemetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.

OS Homo sapiens.

PN WO200153312-A1.

XX 26-JUL-2001.

PF 26-DEC-2000; 2000MO-US34263.

XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0683036.
PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

DR WPI: 2001-442253/47.
N-PSDB: AA161349.

XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 2; SEQ ID NO 7124; 10078bp: English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

CC utilisation of the activities such as: Immune system suppression,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 126 AA:

Query Match 88.9%; Score 32; DB 22; Length 126;
Best Local Similarity 62.5%; Pred. No. 19;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsyxg 8
|| || |
DB 73 wkqsyag 80

RESULT 8
AAE04228
ID AAE04228 standard; Protein: 140 AA.

AC AAE04228;

DT 09-AUG-2001 (first entry)

DE Human gene 10 encoded secreted protein HMP183, SEQ ID NO:83.

XX Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haemopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiogenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;
KW endocrine disorder; infection; wound healing; vulnetary;
KW cell culture; chemotaxis; food additive;
KW binding partner identification.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..24

FT Protein /label= signal_peptide 25..140

FT Misc-difference 136 /note= "Mature secreted protein"

FT WO200136432-A2.

XX 25-MAY-2001.

XX 15-NOV-2000; 2000MO-US31162.

XX 19-NOV-1999; 99US-0166415.

XX 30-JUN-2000; 2000US-0215136.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Komatsoulis GA, Baker KP, Young PE;

XX WPI: 2001-343793/36.
N-PSDB: AAD08518.

XX Isolated nucleic acid molecule encoding a human secreted protein is
PT used in preventing, treating or ameliorating a medical condition -
XX
XX Claim 11; Page 459; 509pp; English.

CC AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted
 CC protein genes, and AAE04199-AAE04239 represent the proteins they encode.
 CC AAE04240-AAE04297 represent human secreted protein fragments or variants.
 CC The secreted proteins and their genes are useful for preventing,
 CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the
 CC amount of the new protein in a sample or by determining the presence of
 CC mutations in the new genes. Specific uses are described for each of the
 CC 18 genes, based on the tissues in which they are most highly expressed,
 CC and include developing products for the diagnosis or treatment of
 CC proliferative disorders, cancer, tumours, foetal and developmental
 CC abnormalities, haematopoietic disorders, diseases of the immune system,
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
 CC allergies, neurological disorders (e.g., Alzheimer's disease,
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
 CC cardiovascular disorders, angiogenic disorders, kidney disorders,
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine
 CC disorders, and infections. The proteins can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin ageing due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues, to identify their
 CC cognate ligands or binding partners, and in chemotaxis, and can be used
 CC as a food additive or preservative to modify storage properties.
 CC Antibodies specific for a protein of the invention can be used in
 CC alleviating symptoms associated with the disorders mentioned above, and
 CC in diagnostic immunoassays (e.g., radioimmunoassay or enzyme linked
 CC immunosorbent assay (ELISA)). The present sequence represents a human
 CC secreted protein of the invention.

XX Sequence 140 AA:

Query Match 88.9%; Score 32; DB 22; Length 140;
 Best Local Similarity 62.5%; Pred. No. 21;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsyxg 8
 || || |
 Db 30 wktssyvg 37

RESULT 9

AAB54052

ID AAB54052 standard; Protein; 155 AA.

XX AAB54052;

DT 09-MAR-2001 (first entry)

DE Human pancreatic cancer antigen protein sequence SEQ ID NO:504.

XX Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
 KW detection; diagnosis; identification; cytostatic; neuroprotective;
 KW neotropic; immunomodulatory; relaxant; contraceptive; gynaecological;
 KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
 KW linkage analysis; tissue identification; tissue typing; forensic;
 KW neural; immune system; muscular; reproductive; gastrointestinal;
 KW pulmonary; cardiovascular; renal; proliferative.

XX Homo sapiens.

PN WO200055320-A1.

PD 21-SEP-2000.

PF 08-MAR-2000; 2000WO-US05989.

PR 12-MAR-1999; 99US-0124270.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;

XX WPI: 2000-579444/54.
 DR N-PSDB; AAC98817.

PT New nucleic acid that is a pancreatic cancer antigen for preventing,
 PT treating, or ameliorating a medical condition, particular pancreatic
 PT cancer, or for use in assays for diagnosing a pathological condition -

PS Claim 11; Page 942-943; 1379pp; English.

XX AAC98773 to AAC99231 encode the human pancreatic cancer associated
 CC proteins, called pancreatic cancer antigens, given in AAB54008 to
 CC AAB54466. The human pancreatic cancer antigens have cytostatic,
 CC neuroprotective, neotropic, immunomodulatory, relaxant, contraceptive,
 CC gynaecological, cardiant and antiinflammatory activities, and can be used
 CC in gene therapy. The polynucleotide and proteins can be used for
 CC preventing, treating, or ameliorating a medical condition or in assays
 CC for diagnosing a pathological condition or a susceptibility to one in a
 CC subject. Binding partners to the proteins and the activity of the
 CC proteins can be identified. The pancreatic cancer antigens can be used to
 CC detect, treat or prevent pancreatic disorders, especially cancer.
 CC Agonists and antagonists to the antigens can be screened for. The
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic
 CC acid hybridisation probes that can be used in chromosome mapping, linkage
 CC analysis, tissue identification and/or typing and a variety of forensic
 CC and diagnostic methods. The proteins can be used to generate antibodies
 CC which are used to purify, detect and target the polypeptides, including
 CC both in vivo and in vitro diagnostic and therapeutic methods. The
 CC proteins can be used to treat or prevent neural, immune system, muscular,
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
 CC sequences used in the exemplification of the present invention.

XX Sequence 155 AA:

Query Match 88.9%; Score 32; DB 21; Length 155;
 Best Local Similarity 62.5%; Pred. No. 23;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsyxg 8
 || || |
 Db 55 wktssyvg 62

RESULT 10

AAY36134

ID AAY36134 standard; Protein; 230 AA.

XX AAY36134;

DT 23-SEP-1999 (first entry)

DE Human secreted protein #6.

XX Secreted protein; human; cytostatic; thrombotic; osteopathic; forensic;
 KW diagnostic; gene therapy; chromosome mapping; secretion vector.

XX Homo sapiens.

PN WO9925825-A2.

PD 27-MAY-1999.

PF 13-NOV-1998; 98WO-IB01862.

PR 04-SEP-1998; 98US-0099273.

PR 13-NOV-1997; 97US-0066677.

PR 17-DEC-1997; 97US-0069957.

PR 09-FEB-1998; 98US-0074121.

PR 13-APR-1998; 98US-0081563.

PR 10-AUG-1998; 98US-0096116.

PA (GEST) GENSET.
XX Bougueleret L, Duclet A, Dumas Milne Edwards J;
XX WPI: 1999-347472/29.
DR N-PSDB; AAX97818.
XX
PT Extended cDNAs encoding secreted proteins
PS Example 28; Page 234-235; 307pp; English.
XX
CC AAY36129-Y36222 represent novel human secreted proteins encoded by the
CC extended cDNA sequences represented in AAX97813-X97906. The proteins
CC of the invention have cytosolic, thrombotic and osteopathic activity.
CC The extended cDNAs can be used to express secreted proteins or parts of
CC them or to obtain antibodies capable of binding to the secreted
CC proteins. They may also be used in diagnostic, forensic, gene therapy
CC and chromosome mapping procedures. Uses also include design of
CC expression vectors and secretion vectors.
XX
SQ Sequence 230 AA:

Query Match 88.9%; Score 32; DB 20; Length 230;
Best Local Similarity 62.5%; Pred. No. 34;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Oy 1 wkxxsyxg 8
|| || |
Db 30 wktsyvg 37

RESULT 11
AAY36181
ID AAY36181 standard; protein: 230 AA.
XX
AC AAY36181;
XX
DT 23-SEP-1999 (first entry)
XX
DE Human secreted protein #53.
XX
KW Secreted protein; human; cytosolic; thrombotic; osteopathic; forensic;
KW diagnostic; gene therapy; chromosome mapping; secretion vector.
XX
OS Homo sapiens.
XX
XX WO9925825-A2.
XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-IB01862.
XX
PR 04-SEP-1998; 98US-0099273.
PR 13-NOV-1997; 97US-006677.
PR 17-DEC-1997; 97US-0069957.
PR 09-FEB-1998; 98US-0074121.
PR 13-APR-1998; 98US-0081563.
PR 10-AUG-1998; 98US-0096116.
XX
PA (GEST) GENSET.
XX
PI Bougueleret L, Duclet A, Dumas Milne Edwards J;
XX
DR WPI: 1999-347472/29.
DR N-PSDB; AAX97865.
XX
PT Extended cDNAs encoding secreted proteins
PS Claim 7; Page 289; 307pp; English.
XX
XX AAY36129-Y36222 represent novel human secreted proteins encoded by the
CC extended cDNA sequences represented in AAX97813-X97906. The proteins

CC of the invention have cytosolic, thrombotic and osteopathic activity.
CC The extended cDNAs can be used to express secreted proteins or parts of
CC them or to obtain antibodies capable of binding to the secreted
CC proteins. They may also be used in diagnostic, forensic, gene therapy
CC and chromosome mapping procedures. Uses also include design of
CC expression vectors and secretion vectors.
XX
SQ Sequence 230 AA:

Query Match 88.9%; Score 32; DB 20; Length 230;
Best Local Similarity 62.5%; Pred. No. 34;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Oy 1 wkxxsyxg 8
|| || |
Db 30 wktsyvg 37

RESULT 12
AAY93378
ID AAY93378 standard; protein: 230 AA.
XX
AC AAY93378;
XX
DT 08-AUG-2000 (first entry)
XX
DE Human PRO1356 (UNQ705) amino acid sequence SEQ ID NO:134.
XX
KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW transmembrane; secretion; immunoadhesion; pharmaceutical; screening.
XX
OS Homo sapiens.
XX
XX WO200012708-A2.
XX
PD 09-MAR-2000.
XX
PF 01-SEP-1999; 99WO-US20111.
XX
PR 01-SEP-1998; 98US-0098716.
PR 01-SEP-1998; 98US-0098749.
PR 01-SEP-1998; 98US-0098750.
PR 02-SEP-1998; 98US-0098803.
PR 02-SEP-1998; 98US-0098821.
PR 02-SEP-1998; 98US-0098843.
PR 09-SEP-1998; 98US-0099536.
PR 09-SEP-1998; 98US-0099598.
PR 09-SEP-1998; 98US-0099602.
PR 09-SEP-1998; 98US-0099642.
PR 10-SEP-1998; 98US-0099741.
PR 10-SEP-1998; 98US-0099754.
PR 10-SEP-1998; 98US-0099792.
PR 10-SEP-1998; 98US-0099793.
PR 10-SEP-1998; 98US-0099808.
PR 10-SEP-1998; 98US-0099812.
PR 10-SEP-1998; 98US-0099815.
PR 10-SEP-1998; 98US-0099816.
PR 15-SEP-1998; 98US-0100385.
PR 15-SEP-1998; 98US-0100388.
PR 15-SEP-1998; 98US-0100390.
PR 16-SEP-1998; 98US-0100584.
PR 16-SEP-1998; 98US-0100627.
PR 16-SEP-1998; 98US-0100661.
PR 16-SEP-1998; 98US-0100662.
PR 16-SEP-1998; 98US-0100664.
PR 17-SEP-1998; 98US-0100683.
PR 17-SEP-1998; 98US-0100684.
PR 17-SEP-1998; 98US-0100710.
PR 17-SEP-1998; 98US-0100711.
PR 17-SEP-1998; 98US-0100919.
PR 17-SEP-1998; 98US-0100930.

PR 18-SEP-1998; 98US-0100848.
 PR 18-SEP-1998; 98US-0100849.
 PR 18-SEP-1998; 98US-0101014.
 PR 18-SEP-1998; 98US-0101068.
 PR 18-SEP-1998; 98US-0101071.
 PR 22-SEP-1998; 98US-0101279.
 PR 23-SEP-1998; 98US-0101471.
 PR 23-SEP-1998; 98US-0101472.
 PR 23-SEP-1998; 98US-0101474.
 PR 23-SEP-1998; 98US-0101475.
 PR 23-SEP-1998; 98US-0101476.
 PR 23-SEP-1998; 98US-0101477.
 PR 23-SEP-1998; 98US-0101479.
 PR 24-SEP-1998; 98US-0101738.
 PR 24-SEP-1998; 98US-0101741.
 PR 24-SEP-1998; 98US-0101743.
 PR 24-SEP-1998; 98US-0101915.
 PR 24-SEP-1998; 98US-0101916.
 PR 29-SEP-1998; 98US-0102207.
 PR 29-SEP-1998; 98US-0102240.
 PR 29-SEP-1998; 98US-0102307.
 PR 29-SEP-1998; 98US-0102331.
 PR 29-SEP-1998; 98US-0102331.
 PR 30-SEP-1998; 98US-0102484.
 PR 30-SEP-1998; 98US-0102487.
 PR 30-SEP-1998; 98US-0102570.
 PR 30-SEP-1998; 98US-0102571.
 PR 01-OCT-1998; 98US-0102684.
 PR 01-OCT-1998; 98US-0102687.
 PR 02-OCT-1998; 98US-0102965.
 PR 06-OCT-1998; 98US-0103258.
 PR 06-OCT-1998; 98US-0103449.
 PR 07-OCT-1998; 98US-0103314.
 PR 07-OCT-1998; 98US-0103315.
 PR 07-OCT-1998; 98US-0103328.
 PR 07-OCT-1998; 98US-0103395.
 PR 07-OCT-1998; 98US-0103396.
 PR 07-OCT-1998; 98US-0103401.
 PR 08-OCT-1998; 98US-0103633.
 PR 08-OCT-1998; 98US-0103679.
 PR 08-OCT-1998; 98US-0103711.
 PR 14-OCT-1998; 98US-0104257.
 PR 20-OCT-1998; 98US-0104987.
 PR 20-OCT-1998; 98US-0105000.
 PR 20-OCT-1998; 98US-0105002.
 PR 21-OCT-1998; 98US-0105104.
 PR 22-OCT-1998; 98US-0105169.
 PR 22-OCT-1998; 98US-0105266.
 PR 26-OCT-1998; 98US-0105693.
 PR 26-OCT-1998; 98US-0105694.
 PR 27-OCT-1998; 98US-0105807.
 PR 27-OCT-1998; 98US-0105881.
 PR 27-OCT-1998; 98US-0105882.
 PR 27-OCT-1998; 98US-0106062.
 PR 28-OCT-1998; 98US-0106023.
 PR 28-OCT-1998; 98US-0106029.
 PR 28-OCT-1998; 98US-0106030.
 PR 28-OCT-1998; 98US-0106032.
 PR 28-OCT-1998; 98US-0106033.
 PR 28-OCT-1998; 98US-0106178.
 PR 29-OCT-1998; 98US-0106248.
 PR 29-OCT-1998; 98US-0106384.
 PR 29-OCT-1998; 98US-0108500.
 PR 30-OCT-1998; 98US-0108464.
 PR 03-NOV-1998; 98US-0106856.
 PR 03-NOV-1998; 98US-0106902.
 PR 03-NOV-1998; 98US-0106905.
 PR 03-NOV-1998; 98US-0106919.
 PR 03-NOV-1998; 98US-0106932.
 PR 03-NOV-1998; 98US-0106934.
 PR 10-NOV-1998; 98US-0107783.
 PR 17-NOV-1998; 98US-0108775.

PR 17-NOV-1998; 98US-0108779.
 PR 17-NOV-1998; 98US-0108787.
 PR 17-NOV-1998; 98US-0108788.
 PR 17-NOV-1998; 98US-0108801.
 PR 17-NOV-1998; 98US-0108802.
 PR 17-NOV-1998; 98US-0108806.
 PR 17-NOV-1998; 98US-0108807.
 PR 17-NOV-1998; 98US-0108867.
 PR 17-NOV-1998; 98US-0108925.
 PR 18-NOV-1998; 98US-0108848.
 PR 18-NOV-1998; 98US-0108849.
 PR 18-NOV-1998; 98US-0108850.
 PR 18-NOV-1998; 98US-0108851.
 PR 18-NOV-1998; 98US-0108852.
 PR 18-NOV-1998; 98US-0108858.
 PR 18-NOV-1998; 98US-0108904.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
 XX
 DR WPI: 2000-237871/20.
 DR N-PSDB: AAA37060.
 XX
 PT New mammalian DNA sequences encoding transmembrane, receptor or
 PT secreted PRO polypeptides, useful for screening of potential peptide or
 PT small molecule inhibitors of the relevant receptor/ligand interactions
 XX
 PS Claim 12; Fig 78; 773pp; English.
 XX
 CC AAA37022 to AAA37144 encode the new isolated human transmembrane,
 CC receptor or secreted PRO polypeptides given in AA99340 to AA99462. The
 CC transmembrane and receptor PRO proteins can be used for screening of
 CC potential peptide or small molecule inhibitors of the relevant
 CC receptor/ligand interactions. The polypeptides and nucleotide sequences
 CC encoding then have various industrial applications, including uses as
 CC pharmaceutical and diagnostic agents. AAA37145 to AAA3730 represent
 CC PCR primers and hybridisation probes used in the isolation of the PRO
 CC polypeptides from the present invention.
 XX
 SQ Sequence 230 AA;
 OY 1 wxxxxxyg 8
 DB 30 wktssyvg 37
 RESULT 13
 ID AA984609 standard; Protein: 230 AA.
 AC AA984609;
 XX
 DT 25-JUL-2000 (First entry)
 XX
 DE A human membrane associated organizational protein (HUNCT).
 XX
 KW Human: membrane associated organizational protein; HUNCT;
 KW cell proliferative disorder; cancer; autoimmune disorder;
 KW inflammatory disorder; neurological disorder; developmental disorder;
 KW vesicle trafficking; reproductive disorder; gastrointestinal disorder;
 KW renal disorder; atherosclerosis; leukaemia; rheumatoid arthritis;
 KW Alzheimer's disease; anxiety; diabetes; ovulatory defect; renal failure;
 KW irritable bowel syndrome; allergy.
 XX
 OS Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers

FT	Modified-site	29	/note- "potential phosphorylation site"
FT	Modified-site	62	/note- "potential phosphorylation site"
FT	Domain	117..138	/note- "transmembrane domain"
FT	Modified-site	155	/note- "potential phosphorylation site"
FT	Domain	164..182	/note- "transmembrane domain"
FT	Modified-site	187	/note- "potential phosphorylation site"
FT	Modified-site	190	/note- "potential glycosylation site"
FT	Modified-site	208	/note- "potential phosphorylation site"
FT	Modified-site	224	/note- "potential phosphorylation site"
XX	WO20018915-A2.		
XX	06-APR-2000.		
XX	23-SEP-1999:	99WO-US22082.	
XX	25-SEP-1998:	98US-0155215.	
PR	13-OCT-1998:	98US-0155251.	
PR	04-MAY-1999:	99US-0172228.	
XX	(INCYTE) INCYTE PHARM INC.		
PI	Yue H, Lal P, Corley NC, Guegler KJ, Baughn MR, Lu AD, Tang YT,		
XX	WPI: 2000-293154/25.		
DR	N-PSDB; AAI12585.		
XX	Human membrane associated organizational protein and nucleic acid		
PT	sequences useful in the diagnosis, treatment and prevention of cell		
PT	proliferative associated disorders e.g. cancer, rheumatoid arthritis		
PT	and Alzheimer's disease -		
XX	Claim 1: Page 76-77; 84pp: English.		
XX	The present sequence represents a membrane associated organizational		
CC	protein (HUNCT). HUNCT is used for the diagnosis, treatment and		
CC	prevention of cell proliferative disorders including cancer and		
CC	autoimmune/inflammatory, neurological, developmental, vesicle		
CC	trafficking, reproductive, gastrointestinal and renal disorders. These		
CC	disorders may include atherosclerosis, leukemia, allergies, rheumatoid		
CC	arthritis, Alzheimer's disease, anxiety, diabetes, ovulatory defects,		
CC	renal failure and irritable bowel syndrome. A vector expressing HUNCT,		
CC	and an agonist of HUNCT can be used to treat or prevent a disorder		
CC	associated with decreased expression or activity of HUNCT. An antagonist		
CC	of HUNCT or a vector expressing the complement of a polynucleotide		
CC	encoding HUNCT can be used to treat or prevent a disorder associated		
CC	with increased expression or activity of HUNCT. Antibodies which bind		
CC	HUNCT can be used for diagnosis of disorders associated with HUNCT		
CC	expression or to monitor patients being treated with HUNCT, agonists,		
CC	antagonists or inhibitors of HUNCT. Assays are preferably carried out on		
CC	body fluids from a patient using radioimmunoassay, enzyme linked		
CC	immunosorbent assays or fluorescent activated cell sorting assays.		
CC	Polynucleotides encoding HUNCT are also used in hybridisation assays to		
CC	determine absence, presence or excess expression of HUNCT and to monitor		
CC	regulation of HUNCT levels during disease therapy.		
XX	Sequence 230 AA:		
SO			
Query Match 88.9%; Score 32; DB 21; Length 230;			
Best Local Similarity 62.5%; Pred. No. 34;			
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;			
OY 1 wxxxxxxg 8			

Db	30 wkssywg 37		
RESULT 14			
AAAM38857			
ID	AAAM38857 standard; Protein: 230 AA.		
XX	AAAM38857;		
AC	22-OCT-2001 (first entry)		
XX			
DT	Human polypeptide SEQ ID NO 2002.		
XX			
DE	Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;		
XX	peripheral nervous system; neuropathy; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		
KW	chemokine; thrombolytic; drug screening; arthritis; inflammation;		
KW	leukemia.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200153312-A1.		
XX	26-JUL-2001.		
PD			
XX	26-DEC-2000; 2000WO-US34263.		
PF			
XX	21-JAN-2000; 2000US-0488725.		
PR	25-APR-2000; 2000US-0552317.		
PR	09-JUL-2000; 2000US-0598042.		
PR	19-JUL-2000; 2000US-0620312.		
PR	03-AUG-2000; 2000US-0653450.		
PR	14-SEP-2000; 2000US-0662191.		
PR	19-OCT-2000; 2000US-0693036.		
PR	29-NOV-2000; 2000US-0727344.		
XX			
PA	(HXSE-) HXSEQ INC.		
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;		
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;		
PI	Zhao QH, Zhou P, Goodrich R, Drmanac RT;		
XX			
DR	WPI: 2001-442253/47.		
DR	N-PSDB; AAI58013.		
XX	Novel nucleic acids and polypeptides, useful for treating disorders		
PT	such as central nervous system injuries -		
XX			
PS	Example 3; SEQ ID NO 2002; 10078pp: English.		
XX			
CC	The invention relates to human nucleic acids (AA157798-AA161369) and		
CC	the encoded polypeptides (AAAM38642-AAAM42213) with neotropic,		
CC	immunosuppressant and cyrostatic activity. The polynucleotides are useful		
CC	in gene therapy. A composition containing a polypeptide or polynucleotide		
CC	of the invention may be used to treat diseases of the peripheral nervous		
CC	system, such as peripheral nervous injuries, peripheral neuropathy and		
CC	localised neuropathies and central nervous system diseases, such as		
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic		
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the		
CC	utilisation of the activities such as: Immune system suppression,		
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, hemostatic		
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,		
CC	assays for receptor activity, arthritis and inflammation, leukemias and		
CC	C.N.S disorders.		
CC	Note: The sequence data for this patent did not form part of the printed		
CC	specification.		
XX	Sequence 230 AA:		
SO			
Query Match 88.9%; Score 32; DB 22; Length 230;			

Best Local Similarity 62.5%; Pred. No. 34;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsyxg 8
11111
Db 30 wktsyvg 37

RESULT 15

AAU12417
ID AAU12417 standard; Protein: 230 AA.

AAU12417;

24-OCT-2001 (first entry)

Human PRO1356 polypeptide sequence.

Human secretory and transmembrane; PRO; mammalian; cancer; lung;
breast; prostate; cervical; tumour necrosis factor- α lpha; TNF- α lpha;
cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
adipocyte; A-peptide; factor VIIA; gene therapy.

Homo sapiens.

WO200140466-A2.

07-JUN-2001.

01-DEC-2000; 2000WO-US32678.

01-DEC-1999; 99WO-US28301.
01-DEC-1999; 99WO-US28634.
02-DEC-1999; 99WO-US28551.
02-DEC-1999; 99WO-US28564.
02-DEC-1999; 99WO-US28565.
09-DEC-1999; 99US-0170262.
16-DEC-1999; 99WO-US30095.
20-DEC-1999; 99WO-US30911.
30-DEC-1999; 99WO-US30999.
06-JAN-2000; 2000WO-US31243.
06-JAN-2000; 2000WO-US00277.
06-JAN-2000; 2000WO-US00376.
11-FEB-2000; 2000WO-US03565.
18-FEB-2000; 2000WO-US04341.
22-FEB-2000; 2000WO-US04414.
24-FEB-2000; 2000WO-US04914.
24-FEB-2000; 2000WO-US05004.
01-MAR-2000; 2000WO-US05601.
20-MAR-2000; 2000WO-US07377.
21-MAR-2000; 2000WO-US07532.
30-MAR-2000; 2000WO-US08439.
17-MAY-2000; 2000WO-US13705.
22-MAY-2000; 2000WO-US14042.
30-MAY-2000; 2000WO-US14941.
02-JUN-2000; 2000WO-US15264.
10-NOV-2000; 2000WO-US30873.

(GETH) GENENTECH INC.

Baker KP, Beresini M, DeForge L, Desnoyers L, Filvaroff E, Gao W;
Gerlitsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
Smith V, Stewart TA, Tumas D, Watanabe CX, Wood WI, Zhang Z;

WPI: 2001-408281/43.

N-PSDE; AAS21489.

isolated, secretory and transmembrane PRO polypeptide used to detect
other PRO polypeptides, link bioactive molecules to cells expressing
PRO polypeptides, and detect the presence of mammalian tumours e.g.
lung, breast, prostate, cervical

PS Claim 12; Fig 492; 813pp; English.

AAU12172-AAU12446 represent novel human secretory and transmembrane
PRO polypeptides. The PRO polypeptides are useful to detect other
PRO polypeptides, to link bioactive molecules to cells expressing
PRO polypeptides, to modulate biological activities of cells expressing
PRO polypeptides, and to detect the presence of mammalian lung, colon,
breast, prostate, rectal, cervical or liver tumours by comparing PRO
polypeptide expression in a cell sample to that in a control sample.
Some of the 275 sequences are also useful to stimulate the release of
tumour necrosis factor- α lpha (TNF- α lpha) from human blood, the
proliferation or differentiation of chondrocytes, the proliferation or
gene expression in pericyte cells, the release of proteoglycans from
cartilage, the proliferation of inner ear utricular supporting cells or
of T-lymphocytes, the release of a cytokine from peripheral blood
monocytes (PBMCs), or the proliferation of endothelial cells. Some of
the PRO polypeptides may modulate glucose or free fatty acid uptake by
skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
to factor VIIA. The PRO polypeptides can be used in assays to identify
CC molecules involved in binding interactions. The polynucleotides encoding
PRO polypeptides can be used to generate probes, antisense RNA/DNA,
transgenic or knock out animals and can be used in gene therapy.

Sequence 230 AA;

Query Match 88.9%; Score 32; DB 22; Length 230;

Best Local Similarity 62.5%; Pred. No. 34;

Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsyxg 8
11111
Db 30 wktsyvg 37

RESULT 16

AAE04207
ID AAE04207 standard; Protein: 230 AA.

AAE04207;

09-AUG-2001 (first entry)

Human gene 10 encoded secreted protein HTP1H83, SEQ ID NO:62.

Human; secreted protein; proliferative disorder; cancer; tumour;
foetal abnormality; developmental abnormality; haematopoietic disorder;
immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
inflammation; allergy; neurological disorder; Alzheimer's disease;
Parkinson's disease; cognitive disorder; schizophrenia; asthma;
skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
cardiovascular disorder; angiogenic disorder; kidney disorder;
gastrointestinal disorder; pregnancy-related disorder; gene therapy;
endocrine disorder; infection; wound healing; vulnerability;
cell culture; chemotaxis; food additive;
binding partner identification.

Homo sapiens.

Key Location/Qualifiers

Peptide 1..24

Protein /label= signal_peptide

FT /note= "Mature secreted protein"

WO200136432-A2.

25-MAY-2001.

15-NOV-2000; 2000WO-US31162.

19-NOV-1999; 99US-0166415.

30-JUN-2000; 2000US-0215136.

XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Ruben SM, Komatsoulis GA, Baker KP, Young PE;
XX DR MPI: 2001-243793/36.
XX N-PSTDB: AAD08497.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is
XX used in preventing, treating or ameliorating a medical condition -
XX
XX Claim 11; Page 439; 509pp: English.

CC AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted
CC protein genes, and AAEO4199-AAEO4239 represent the proteins they encode.
CC AAEO4240-AAEO4297 represent human secreted protein fragments or variants.
CC The secreted proteins and their genes are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 18 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., Rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin ageing due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein of the invention.

XX
XX SQ Sequence 230 AA;

Query Match 88.9%; Score 32; DB 22; Length 230;
Best Local Similarity 62.5%; Pred No. 34;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 WXXXXXGXG 8
|| || |
Db 30 WKTSSYG 37

RESULT 17
AAB87565
ID AAB87565 standard; Protein: 230 AA.
XX
XX AAB87565;
AC
XX
XX 15-MAY-2001 (first entry)
XX
XX DE Human PRO1356.
XX
XX Human; PRO protein; mapping.
XX
XX Homo sapiens.
XX OS
XX PN WO200116318-A2.
XX
XX PD 08-MAR-2001.

```

FE 24-AUG-2000; 2000MO-US23328.
XX
PR 01-SEP-1999; 99MO-US20111.
PR 15-SEP-1999; 99MO-US21090.
PR 07-DEC-1999; 99US-0169495.
PR 09-DEC-1999; 99US-0170262.
PR 11-JAN-2000; 2000US-0175481.
PR 18-FEB-2000; 2000MO-US04341.
PR 18-FEB-2000; 2000MO-US04342.
PR 22-FEB-2000; 2000MO-US04414.
PR 01-MAR-2000; 2000MO-US05601.
PR 03-MAR-2000; 2000US-0187202.
PR 25-APR-2000; 2000US-0199397.
PR 22-MAY-2000; 2000MO-US14042.
PR 05-JUN-2000; 2000US-0209832.
XX
PA (GETH ) GENENTECH INC.
PI Eaton DL, Filyavroff E, Gerritsen ME, Goddard A, Godowski PJ;
PI Girmaldi CJ, Gurney AL, Watanabe CK, Wood WI;
XX
DR WPI: 2001-183260/18.
DR N-PSDB; AAF92097.
XX
XX Eighty four nucleic acids encoding PRO polypeptides, useful in
PT molecular biology, including use as hybridization probes, and in
PT chromosome and gene mapping.
XX
PS Claim 12; Fig 80; 278pp; English.
XX
CC The present sequence is a human PRO polypeptide (secreted and
CC transmembrane). The PRO protein, and PRO agonists, PRO antagonists or
CC anti-PRO antibodies are useful for preparation of a medicament useful in
CC the treatment of a condition which is responsive to the PRO protein,
CC agonists, antagonists or anti-PRO antibodies. The PRO protein may also be
CC employed as molecular weight markers for protein electrophoresis. The PRO
CC coding sequence has applications in molecular biology, including use as
CC hybridisation probes, and in chromosome and gene mapping.
XX
SQ Sequence 230 AA;
XX
OY 1 wxsxyxg 8
|||
30 wktssyvg 37
DB
RESULT 18
AAB88342
IID AAB88342 standard; Protein; 230 AA.
XX
AC AAB88342;
XX
DT 23-MAY-2001 (first entry)
XX
DE Human membrane or secretory protein clone PSEC0059.
XX
KW Human; secretory protein; membrane protein; vaccine; gene therapy;
KW rheumatoid arthritis; diabetes.
XX
OS Homo sapiens.
XX
PN EP1067182-A2.
XX
PD 10-JAN-2001.
XX
PF 07-JUL-2000; 2000EP-0114090.
XX
PR 08-JUL-1999; 99JP-0194179.

```

```

PR 11-JAN-2000; 2000JP-0118775.
PR 02-MAY-2000; 2000JP-0183766.
PA (HELI-) HELIX RES INSTR.
PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
DR WPI; 2001-093989/11.
DR N-PSDB; AAF93769.
PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT gene therapy or as candidate target molecules in drug development -
XX
XX Claim 1; SEQ ID 52; 609pp + CD ROM; English.
XX
XX This invention relates to nucleic acid sequences AAF93744 - AAF93916
XX which encode human secretory or membrane proteins represented by
XX AAF88317 - AAF88419. Included in the invention are primers
XX AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
XX cDNA sequences of the invention. The invention also includes methods for
XX the production of antibodies directed against the proteins, and cDNA
XX sequences, which can be used in vaccines. The polynucleotide sequences
XX can be used in gene therapy. The polynucleotide sequences and the
XX proteins they encode may be used in the prevention, treatment and
XX diagnosis of diseases associated with inappropriate secretory
XX protein/membrane protein expression. The nucleic acids and complementary
XX sequences may also be used as DNA probes in diagnostic assays
XX (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
XX presence of similar nucleic acid sequences in samples. They may also be
XX used to study the expression and function of secretory proteins/membrane
XX polypeptides and their role in metabolism. The polypeptides may be used
XX as antigens in the production of antibodies against them and in assays to
XX identify modulators (agonists and antagonists) of expression and
XX activity. The antibodies and antagonists may also be used as therapeutic
XX agents to down regulate expression and activity. The antibodies may also
XX be used as diagnostic agents for detecting the presence of the
XX polypeptides in samples (e.g. by enzyme linked immunosorbant assay
XX (ELISA). Examples of diseases which may be treated include rheumatoid
XX arthritis and diabetes.
XX
XX Sequence 230 AA:
SO

```

```

Query Match 88.9%; Score 32; DB 22; Length 230;
Best Local Similarity 62.5%; Pred. No. 34;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 wkxxsyxg 8
   || || |
Db 30 wktssyvg 37

```

```

RESULT 19
AAB66127
ID AAB66127 standard; protein; 230 AA.
XX
XX AAB66127;
AC
XX 02-APR-2001 (first entry)
DT
XX Protein of the invention #39.
DE
XX
XX Secreted; transmembrane; gene therapy.
XX
XX Unidentified.
XX
XX WO200078961-A1.
XX
XX 28-DEC-2000.
XX
XX 18-FEB-2000; 2000WO-US04342.
XX
XX 23-JUN-1999; 99US-0141037.
PR

```

```

PR 20-JUL-1999; 99US-0144758.
PR 26-JUL-1999; 99US-0145698.
PR 01-SEP-1999; 99WO-US20111.
PR 29-OCT-1999; 99US-0162506.
PR 30-NOV-1999; 99WO-US28313.
PR 02-DEC-1999; 99WO-US28551.
PR 16-DEC-1999; 99WO-US30095.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00376.
XX
XX (GENTH ) GENENTECH INC.
XX
XX Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;
XX Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
XX Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;
XX Watanabe CK, Williams PM, Wood WI;
XX WPI; 2001-071395/08.
XX
XX Secreted and transmembrane proteins and nucleic acids designated PRO,
XX useful as hybridization probes, in chromosome and gene mapping and gene
XX therapy -
XX
XX Claim 1; Fig 78; 787pp; English.
XX
XX The present invention relates to secreted and transmembrane proteins.
XX These proteins and the DNA encoding them may be used as hybridization
XX probes, in chromosome and gene mapping and in the generation of
XX anti-sense RNA and DNA. They may also be used to generate either
XX transgenic animals or knockout animals which are in turn useful for
XX development and screening of therapeutically useful reagents.
XX The nucleic acids may also be used in gene therapy.
XX
XX Sequence 230 AA:
SO

```

```

Query Match 88.9%; Score 32; DB 22; Length 230;
Best Local Similarity 62.5%; Pred. No. 34;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 wkxxsyxg 8
   || || |
Db 30 wktssyvg 37

```

```

RESULT 20
AAM40643
ID AAM40643 standard; protein; 260 AA.
XX
XX AAM40643;
AC
XX
XX 22-OCT-2001 (first entry)
DT
XX
XX Human polypeptide SEQ ID NO 5574.
DE
XX
XX Human;
XX
XX Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia.
XX
XX Homo sapiens.
XX
XX
XX WO200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-048725.
XX
XX 25-APR-2000; 2000US-0552317.
XX
XX 09-JUL-2000; 2000US-0598042.
PR

```

PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

PI Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI: 2001-442253/47.
DR N-PSDB: AAI59799.

XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -

PS Example 2; SEQ ID NO 5574; 10078bp; English.

CC The invention relates to human nucleic acids (AAI5798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localized neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX Sequence 260 AA:

Query Match 88.9%; Score 32; DB 22; Length 260;

Best Local Similarity 62.5%; Pred. NO. 38;

Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 WXXXXXXG 8

DB 60 WXXXXXXG 67

Search completed: January 14, 2002, 07:35:58
Job time: 809 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: January 14, 2002, 07:22:29 ; Search time 48.05 Seconds
(without alignments)
3.747 Million cell updates/sec

Title: 09-185908-1a
Perfect score: 36
Sequence: 1 wkxxsyxg 8

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues
Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfilest.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	75.0	344	2 US-07-857-224B-92	Sequence 92, Appl
2	27	75.0	344	2 US-07-857-224B-93	Sequence 93, Appl
3	27	75.0	345	3 US-09-222-817-2	Sequence 2, Appl1
4	27	75.0	345	4 US-09-222-817-2	Sequence 2, Appl1
5	27	75.0	530	3 US-09-222-817-12	Sequence 12, Appl1
6	27	75.0	530	3 US-09-222-817-14	Sequence 14, Appl1
7	27	75.0	530	4 US-09-222-786-12	Sequence 12, Appl1
8	27	75.0	530	4 US-09-222-786-14	Sequence 14, Appl1
9	27	75.0	748	1 US-08-408-318-2	Sequence 2, Appl1
10	27	75.0	748	1 US-08-369-796-10	Sequence 10, Appl1
11	27	75.0	748	1 US-08-839-164-2	Sequence 2, Appl1
12	27	75.0	748	2 US-08-852-091-10	Sequence 10, Appl1
13	27	75.0	748	2 US-08-820-754-10	Sequence 10, Appl1
14	27	75.0	748	3 US-08-956-652-10	Sequence 10, Appl1
15	27	75.0	748	3 US-08-956-869-10	Sequence 10, Appl1
16	27	75.0	748	3 US-08-948-547-10	Sequence 10, Appl1
17	27	75.0	748	4 US-08-087-465-8	Sequence 8, Appl1
18	27	75.0	748	5 PCT-US95-17025-10	Sequence 10, Appl1
19	27	75.0	749	1 US-08-276-099A-15	Sequence 15, Appl1
20	27	75.0	749	1 US-08-781-890-15	Sequence 15, Appl1
21	27	75.0	749	3 US-09-012-710-9	Sequence 9, Appl1
22	27	75.0	749	4 US-09-364-970-4	Sequence 4, Appl1
23	27	75.0	851	1 US-08-369-796-2	Sequence 2, Appl1
24	27	75.0	851	2 US-08-852-091-2	Sequence 2, Appl1
25	27	75.0	851	2 US-08-820-754-2	Sequence 2, Appl1
26	27	75.0	851	3 US-08-956-652-2	Sequence 2, Appl1
27	27	75.0	851	3 US-08-956-869-2	Sequence 2, Appl1

28	27	75.0	851	3 US-09-012-710-2	Sequence 2, Appl1
29	27	75.0	851	3 US-08-948-547-2	Sequence 2, Appl1
30	27	75.0	851	4 US-09-087-465-4	Sequence 4, Appl1
31	27	75.0	851	5 US-09-364-970-2	Sequence 2, Appl1
32	27	75.0	851	5 PCT-US95-17025-2	Sequence 2, Appl1
33	27	75.0	852	1 US-08-276-099A-13	Sequence 13, Appl1
34	27	75.0	852	1 US-08-781-890-13	Sequence 13, Appl1
35	26	72.2	378	2 US-08-244-205-7	Sequence 7, Appl1
36	26	72.2	378	5 PCT-US92-10284-7	Sequence 7, Appl1
37	26	72.2	404	2 US-08-244-205-9	Sequence 9, Appl1
38	26	72.2	404	5 PCT-US92-10284-9	Sequence 9, Appl1
39	26	72.2	435	4 US-09-072-917A-9	Sequence 9, Appl1
40	26	72.2	446	2 US-08-244-205-5	Sequence 5, Appl1
41	26	72.2	446	5 PCT-US92-10284-5	Sequence 5, Appl1
42	26	72.2	446	5 PCT-US94-01321-10	Sequence 10, Appl1
43	26	72.2	483	2 US-08-770-544-8	Sequence 8, Appl1
44	26	72.2	497	1 US-08-276-635B-5	Sequence 5, Appl1
45	26	72.2	497	3 US-08-464-256B-5	Sequence 5, Appl1
46	26	72.2	497	3 US-08-471-961-5	Sequence 5, Appl1
47	26	72.2	971	2 US-08-724-334D-22	Sequence 22, Appl1
48	26	72.2	971	3 US-09-270-984A-22	Sequence 22, Appl1
49	26	72.2	971	3 US-09-177-431-8	Sequence 8, Appl1
50	25	69.4	41	4 US-08-706-344C-23	Sequence 23, Appl1

ALIGNMENTS

RESULT 1
US-07-857-224B-92
; Sequence 92, Application US/07857224B
; Patent No. 5958784
; GENERAL INFORMATION:
; APPLICANT: Benner, Steven A.
; TITLE OF INVENTION: Predicting Folded Structures of Proteins
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven A. Benner
; STREET: Hadlaubstrasse 151
; CITY: Zurich
; STATE: none
; COUNTRY: Switzerland
; ZIP: (note: this is an international post code) CH-8092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/857, 224B
; FILING DATE: 03/25/92
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA: none
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (International) 41 1 632 2830
; TELEFAX: (International) 41 1 262 2437
; TELEX: none
; INFORMATION FOR SEQ ID NO: 92:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 344
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: protein
; ORIGINAL SOURCE:
; ORGANISM: Scizosaccharomyces pombe
; FEATURE: Alcohol dehydrogenase, Table 3 Column 15
; PUBLICATION INFORMATION:
; AUTHORS: Joernvall, H.
; AUTHORS: Persson, M.
; AUTHORS: Jeffery, J.
; TITLE: Alcohol dehydrogenases

JOURNAL: Proceedings of the National Academy of Sciences, USA
VOLUME: 78
PAGES: 4226-4230
DATE: 1981
US-07-857-224B-92

Query Match 75.0%; Score 27; DB 2; Length 344;
Best Local Similarity 50.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 wkxxsyxg 8
| | | |
Db 82 WKIGDYAG 89

RESULT 2
US-07-857-224B-93
; Sequence 93, Application US/07857224B
; Patent No. 5958784
; GENERAL INFORMATION:
; APPLICANT: Benner, Steven A.
; TITLE OF INVENTION: Predicting Folded Structures of Proteins
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven A. Benner
; STREET: Hadlaubstrasse 151
; CITY: Zurich
; STATE: none
; COUNTRY: Switzerland
; ZIP: (note: this is an international post code) CH-8092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/857,224B
; FILING DATE: 03/25/92
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (international) 41 1 632 2830
; TELEFAX: (international) 41 1 262 2437
; TELEX: none
; INFORMATION FOR SEO ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 344
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: protein
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus
; FEATURE: Alcohol dehydrogenase, Table 3 Column 16
; PUBLICATION INFORMATION:
; AUTHORS:
; AUTHORS: Joernvall, H.
; AUTHORS: Persson, M.
; AUTHORS: Jeffery, J.
; TITLE: Alcohol dehydrogenases
; JOURNAL: Proceedings of the National Academy of Sciences, USA
; VOLUME: 78
; PAGES: 4226-4230
; DATE: 1981
; US-07-857-224B-93

Query Match 75.0%; Score 27; DB 2; Length 344;
Best Local Similarity 50.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 wkxxsyxg 8

Db 82 WKIGDYAG 89

RESULT 3
US-09-222-817-2
; Sequence 2, Application US/09222817
; Patent No. 6037154
; GENERAL INFORMATION:
; APPLICANT: MIKIKO SUGA, Masakazu SUGIMOTO, Tsuyoshi OSUMI, Tsuyoshi NAKAMATSU,
; TITLE OF INVENTION: METHOD OF PRODUCING L-SERINE BY FERMENTATION
; FILE REFERENCE: OP813
; CURRENT APPLICATION NUMBER: US/09/222,817
; CURRENT FILING DATE: 1998-12-30
; EARLIER APPLICATION NUMBER: JP 10-3751
; EARLIER FILING DATE: 1998-01-12
; EARLIER APPLICATION NUMBER: JP 10-353521
; EARLIER FILING DATE: 1998-12-11
; NUMBER OF SEO ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEO ID NO 2
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-222-817-2

Query Match 75.0%; Score 27; DB 3; Length 345;
Best Local Similarity 50.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsyxg 8
| | | |
Db 132 WKRSSFNG 139

RESULT 4
US-09-222-786-2
; Sequence 2, Application US/09222786A
; Patent No. 6258573
; GENERAL INFORMATION:
; APPLICANT: MIKIKO SUGA, Masakazu SUGIMOTO, Tsuyoshi OSUMI, Tsuyoshi NAKAMATSU,
; TITLE OF INVENTION: METHOD OF PRODUCING L-SERINE BY FERMENTATION
; FILE REFERENCE: OP812
; CURRENT APPLICATION NUMBER: US/09/222,786A
; CURRENT FILING DATE: 1998-12-30
; EARLIER APPLICATION NUMBER: JP 10-3751
; EARLIER FILING DATE: 1998-01-12
; EARLIER APPLICATION NUMBER: JP 10-353513
; EARLIER FILING DATE: 1998-12-11
; NUMBER OF SEO ID NOS: 14
; SOFTWARE: Patentln Ver. 2.0
; SEO ID NO 2
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-222-786-2

Query Match 75.0%; Score 27; DB 4; Length 345;
Best Local Similarity 50.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsyxg 8
| | | |
Db 132 WKRSSFNG 139

RESULT 5
US-09-222-817-12
; Sequence 12, Application US/09222817
; Patent No. 6037154
; GENERAL INFORMATION:


```

; APPLICANT: MIKIKO SUGA, Masakazu SUGIMOTO, Tsuyoshi OSUMI, Tsuyoshi NAKAMATSU,
; TITLE OF INVENTION: METHOD OF PRODUCING L-SERINE BY FERMENTATION
; FILE REFERENCE: OP813
; CURRENT APPLICATION NUMBER: US/09/222,817
; EARLIER FILING DATE: 1998-12-30
; EARLIER APPLICATION NUMBER: JP 10-3751
; EARLIER FILING DATE: 1998-01-12
; EARLIER APPLICATION NUMBER: JP 10-353521
; EARLIER FILING DATE: 1998-12-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 530
; TYPE: PRF
; ORGANISM: Brevibacterium flavum
US-09-222-817-12

Query Match          75.0%; Score 27; DB 3; Length 530;
Best Local Similarity 50.0%; Pred. No. 2.5e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxyxg 8
   || | : |
   Db 132 WKRSSFNG 139

RESULT 6
US-09-222-817-14
; Sequence 14, Application US/09222817
; Patent No. 6037154
; GENERAL INFORMATION:
; APPLICANT: MIKIKO SUGA, Masakazu SUGIMOTO, Tsuyoshi OSUMI, Tsuyoshi NAKAMATSU,
; TITLE OF INVENTION: METHOD OF PRODUCING L-SERINE BY FERMENTATION
; FILE REFERENCE: OP813
; CURRENT APPLICATION NUMBER: US/09/222,817
; CURRENT FILING DATE: 1998-12-30
; EARLIER APPLICATION NUMBER: JP 10-3751
; EARLIER FILING DATE: 1998-01-12
; EARLIER APPLICATION NUMBER: JP 10-353521
; EARLIER FILING DATE: 1998-12-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 530
; TYPE: PRF
; ORGANISM: Brevibacterium flavum
US-09-222-817-14

Query Match          75.0%; Score 27; DB 3; Length 530;
Best Local Similarity 50.0%; Pred. No. 2.5e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxyxg 8
   || | : |
   Db 132 WKRSSFNG 139

RESULT 7
US-09-222-786-12
; Sequence 12, Application US/09222786A
; Patent No. 6258573
; GENERAL INFORMATION:
; APPLICANT: MIKIKO SUGA, Masakazu SUGIMOTO, Tsuyoshi OSUMI, Tsuyoshi NAKAMATSU,
; TITLE OF INVENTION: METHOD OF PRODUCING L-SERINE BY FERMENTATION
; FILE REFERENCE: OP812
; CURRENT APPLICATION NUMBER: US/09/222,786A
; CURRENT FILING DATE: 1998-12-30
; EARLIER APPLICATION NUMBER: JP 10-3751
; EARLIER FILING DATE: 1998-01-12
; EARLIER APPLICATION NUMBER: JP 10-353513
; EARLIER FILING DATE: 1998-12-11

```

```

; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 530
; TYPE: PRF
; ORGANISM: Brevibacterium flavum
US-09-222-786-12

Query Match          75.0%; Score 27; DB 4; Length 530;
Best Local Similarity 50.0%; Pred. No. 2.5e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxyxg 8
   || | : |
   Db 132 WKRSSFNG 139

RESULT 8
US-09-222-786-14
; Sequence 14, Application US/09222786A
; Patent No. 6258573
; GENERAL INFORMATION:
; APPLICANT: MIKIKO SUGA, Masakazu SUGIMOTO, Tsuyoshi OSUMI, Tsuyoshi NAKAMATSU,
; TITLE OF INVENTION: METHOD OF PRODUCING L-SERINE BY FERMENTATION
; FILE REFERENCE: OP812
; CURRENT APPLICATION NUMBER: US/09/222,786A
; CURRENT FILING DATE: 1998-12-30
; EARLIER APPLICATION NUMBER: JP 10-3751
; EARLIER FILING DATE: 1998-01-12
; EARLIER APPLICATION NUMBER: JP 10-353513
; EARLIER FILING DATE: 1998-12-11
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 530
; TYPE: PRF
; ORGANISM: Brevibacterium flavum
US-09-222-786-14

Query Match          75.0%; Score 27; DB 4; Length 530;
Best Local Similarity 50.0%; Pred. No. 2.5e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxyxg 8
   || | : |
   Db 132 WKRSSFNG 139

RESULT 9
US-08-408-318-2
; Sequence 2, Application US/08408318
; Patent No. 5639858
; GENERAL INFORMATION:
; APPLICANT: Hoey, Timothy
; TITLE OF INVENTION: Human Signal Transducers and Binding
; TITLE OF INVENTION: Assays
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohnbach, Test, Albritton & Herber
; STREET: 850 Hansen Way, #200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/408,318

```

FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-60845
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-494-8700
TELEFAX: 415-494-8771
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 748 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-408-318-2

Query Match 75.0%; Score 27; DB 1; Length 748;
Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsyxg 8
1: 111
Db 500 WOFSSYVG 507

RESULT 10
US-08-369-796-10
Sequence 10, Application US/08369796
Patent No. 5716622
GENERAL INFORMATION:
APPLICANT: James E. Darnell, Jr.
APPLICANT: Zilong Wen
APPLICANT: Curt M. Horvath
APPLICANT: Zhong Zhong
TITLE OF INVENTION: FUNCTIONALLY ACTIVE REGIONS OF SIGNAL
TITLE OF INVENTION: TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION (STAT) PROTEINS
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/369,796
FILING DATE: 06-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-116
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 748 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-369-796-10

Query Match 75.0%; Score 27; DB 1; Length 748;

Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsyxg 8
1: 111
Db 500 WOFSSYVG 507

RESULT 11
US-08-839-164-2
Sequence 2, Application US/08839164
Patent No. 5756700
GENERAL INFORMATION:
APPLICANT: Hoey, Timothy
TITLE OF INVENTION: Human Signal Transducers and Binding
TITLE OF INVENTION: Assays
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herber
STREET: 850 Hansen Way, #200
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,164
FILING DATE: 23-APR-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/408,318
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-60845
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-494-8700
TELEFAX: 415-494-8771
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 748 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-839-164-2

Query Match 75.0%; Score 27; DB 1; Length 748;
Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsyxg 8
1: 111
Db 500 WOFSSYVG 507

RESULT 12
US-08-852-091-10
Sequence 10, Application US/08852091
Patent No. 5863228
GENERAL INFORMATION:
APPLICANT: James E. Darnell, Jr.
APPLICANT: Zilong Wen
APPLICANT: Curt M. Horvath
APPLICANT: Zhong Zhong
TITLE OF INVENTION: FUNCTIONALLY ACTIVE REGIONS OF SIGNAL
TITLE OF INVENTION: TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION (STAT) PROTEINS
NUMBER OF SEQUENCES: 39

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Klauber & Jackson
;; STREET: 411 Hackensack Avenue
;; CITY: Hackensack
;; STATE: New Jersey
;; COUNTRY: USA
;; ZIP: 07601
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/852,091
;; FILING DATE: 06-MAY-1997
;; CLASSIFICATION: 424
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: 08/369,796
;; FILING DATE: 06-JAN-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Jackson Esq., David A.
;; REGISTRATION NUMBER: 26,742
;; REFERENCE/DOCKET NUMBER: 600-1-116
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 201 487-5800
;; TELEFAX: 201 343-1684
;; TELEX: 133521
;; INFORMATION FOR SEQ ID NO: 10:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 748 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-852-091-10

Query Match 75.0%; Score 27; DB 2; Length 748;
Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxyg 8
I: | | |
Db 500 WQFSSYVG 507

RESULT 13
US-08-820-754-10
; Sequence 10, Application US/08820754
; Patent No. 5976835
; GENERAL INFORMATION:
; APPLICANT: Darnell Jr., James E.
; APPLICANT: Schindler, Christian W.
; APPLICANT: Fu, Xian-Yuan
; APPLICANT: Wen, Zilong
; APPLICANT: Zhong, Zhong
; TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN
; TITLE OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/820,754
; FILING DATE: 19-MAR-1997

;; CLASSIFICATION: 530
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: US 08/212,185
;; FILING DATE: 11-MAR-1994
;; APPLICATION NUMBER: US 07/980,498
;; FILING DATE: 23-NOV-1992
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: US 07/854,296
;; FILING DATE: 19-MAR-1992
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: WO US93/02569
;; FILING DATE: 19-MAR-1993
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: US 08/126,588
;; FILING DATE: 24-SEP-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Jackson Esq., David A.
;; REGISTRATION NUMBER: 26,742
;; REFERENCE/DOCKET NUMBER: 600-1-073 CIP
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 201 487-5800
;; TELEFAX: 201 343-1684
;; TELEX: 133521
;; INFORMATION FOR SEQ ID NO: 10:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 748 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-820-754-10

Query Match 75.0%; Score 27; DB 2; Length 748;
Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxyg 8
I: | | |
Db 500 WQFSSYVG 507

RESULT 14
US-08-956-652-10
; Sequence 10, Application US/08956652
; Patent No. 6013475
; GENERAL INFORMATION:
; APPLICANT: Darnell Jr., James E.
; APPLICANT: Schindler, Christian W.
; APPLICANT: Fu, Xian-Yuan
; APPLICANT: Wen, Zilong
; APPLICANT: Zhong, Zhong
; TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN
; TITLE OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,652
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,185
; FILING DATE: 11-MAR-1994

APPLICATION NUMBER: US 07/980,498
FILING DATE: 23-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/854,296
FILING DATE: 19-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO US93/02569
FILING DATE: 19-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/126,588
FILING DATE: 24-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-073 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 748 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-956-652-10

Query Match 75.0%; Score 27; DB 3; Length 748;
Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxyxg 8
1: 111
DB 500 WQFSSYVG 507

RESULT 15
US-08-956-869-10
Sequence 10, Application US/08956869
Patent No. 6030808
GENERAL INFORMATION:
APPLICANT: Darnell Jr., James E.
APPLICANT: Schindler, Christian W.
APPLICANT: Fu, Xian-Yuan
APPLICANT: Wen, Zilong
APPLICANT: Zhong, Zhong
TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN
TITLE OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,869
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/212,185
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/854,296
FILING DATE: 19-MAR-1992
PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO US93/02569
FILING DATE: 19-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/126,588
FILING DATE: 24-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-073 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 748 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-956-869-10

Query Match 75.0%; Score 27; DB 3; Length 748;
Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxyxg 8
1: 111
DB 500 WQFSSYVG 507

RESULT 16
US-08-948-547-10
Sequence 10, Application US/08948547
Patent No. 6124118
GENERAL INFORMATION:
APPLICANT: Darnell Jr., James E.
APPLICANT: Schindler, Christian W.
APPLICANT: Fu, Xian-Yuan
APPLICANT: Wen, Zilong
APPLICANT: Zhong, Zhong
TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN
TITLE OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/948,547
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/212,185
FILING DATE: 11-MAR-1994
APPLICATION NUMBER: US 07/980,498
FILING DATE: 23-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/854,296
FILING DATE: 19-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO US93/02569
FILING DATE: 19-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/126,588

FILING DATE: 24-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: JACKSON ESQ., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-073 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 748 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-948-547-10

Query Match 75.0%; Score 27; DB 3; Length 748;
Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxxxxyxg 8
1: 1111
DB 500 WOFSSYVG 507

RESULT 17
US-09-087-465-8
Sequence 8, Application US/09087465A
Patent No. 6160092
GENERAL INFORMATION:
APPLICANT: Vinkemeler, Uwe
APPLICANT: Chen, Xiaomln
APPLICANT: Darnell Jr., James E
APPLICANT: Kuriyan, John
TITLE OF INVENTION: A CRYSTAL OF THE CORE PORTION OF A STAT AND METHODS OF
TITLE OF INVENTION: USE
FILE REFERENCE: 600-1-229
CURRENT APPLICATION NUMBER: US/09/087,465A
CURRENT FILING DATE: 1998-05-29
NUMBER OF SEQ ID NOS: 37
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 8
LENGTH: 748
TYPE: PRT
ORGANISM: Homo sapiens
US-09-087-465-8

Query Match 75.0%; Score 27; DB 4; Length 748;
Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxxxxyxg 8
1: 1111
DB 500 WOFSSYVG 507

RESULT 18
PCT-US95-17025-10
Sequence 10, Application PC/US9517025
GENERAL INFORMATION:
APPLICANT: James E. Darnell, Jr.
APPLICANT: Zilong Wen
APPLICANT: Curt M. Hoyvalth
APPLICANT: Zhong Zhong
TITLE OF INVENTION: FUNCTIONALLY ACTIVE REGIONS OF SIGNAL
TITLE OF INVENTION: TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION (STAT) PROTEINS
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauder & Jackson
STREET: 411 Hackensack Avenue

CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/17025
FILING DATE: 28-DEC-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/369,796
FILING DATE: 06-JAN-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: JACKSON ESQ., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-116
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 748 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-17025-10

Query Match 75.0%; Score 27; DB 5; Length 748;
Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxxxxyxg 8
1: 1111
DB 500 WOFSSYVG 507

RESULT 19
US-08-276-099A-15
Sequence 15, Application US/08276099A
Patent No. 5591825
GENERAL INFORMATION:
APPLICANT: McKnight, Steven L
APPLICANT: Hou, Jinzhao
TITLE OF INVENTION: INTERLEUKIN-4 SIGNAL TRANSDUCERS AND
TITLE OF INVENTION: BINDING ASSAYS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276,099A
FILING DATE: 15-JUL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard Aron
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59451-1/RAO

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 749 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-276-099A-15

Query Match 75.0%; Score 27; DB 1; Length 749;
Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxyxg 8
1: 111
Db 501 WQFSYVG 508

RESULT 20
US-08-781-890-15
Sequence 15, Application US/08781890
Patent No. 5710266
GENERAL INFORMATION:
APPLICANT: McKnight, Steven L
APPLICANT: Hou, Jinhao
TITLE OF INVENTION: INTERLEUKIN-4 SIGNAL TRANSDUCERS AND
TITLE OF INVENTION: BINDING ASSAYS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,890
FILING DATE: 05-JAN-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/276,099
FILING DATE: 15-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard Atton
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59451-1/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 749 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-781-890-15

Query Match 75.0%; Score 27; DB 1; Length 749;
Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxyxg 8
1: 111
Db 501 WQFSYVG 508

Search completed: January 14, 2002, 07:23:41
Job time: 72 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: January 14, 2002, 07:37:25 ; Search time 63.57 Seconds
(without alignments)
9.586 Million coll updates/sec

Title: 09-185908-lb
Perfect score: 36
Sequence: 1 wxxxxxyxg 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database :
1: p1r1:
2: p1r2:
3: p1r3:
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	83.3	538	2 F83354	probable sulfatase
2	29	80.6	6805	2 S20901	titin - rabbit (fr
3	29	80.6	26926	1 T38344	titin, cardiac mus
4	28	77.8	65	2 S23154	light-harvesting p
5	28	77.8	213	2 A56152	major 25k outer me
6	28	77.8	303	2 J01366	hypothetical 33k p
7	28	77.8	341	2 T48858	acetylpolymyosin am
8	28	77.8	411	2 D64884	conserved hypothet
9	28	77.8	411	2 H85748	probable integrase
10	28	77.8	441	2 S60618	4-alpha-glucanotra
11	28	77.8	513	1 A35742	aqualysin (Ec 3.4.
12	28	77.8	4152	2 T31102	filamentous hemagg
13	28	77.8	4919	2 T31105	hypothetical prote
14	27	75.0	109	2 D82576	hypothetical prote
15	27	75.0	159	1 OOCV2	aphid transmission
16	27	75.0	246	2 T32510	hypothetical prote
17	27	75.0	260	2 F82120	zinc ABC transport
18	27	75.0	261	2 A64066	probable membrane
19	27	75.0	262	2 F82959	permease of ABC z1
20	27	75.0	288	2 D75286	serine proteinase
21	27	75.0	303	2 T00479	probable phosphat
22	27	75.0	368	1 D64763	alcohol dehydrogen
23	27	75.0	369	1 S57325	alcohol dehydrogen
24	27	75.0	369	1 G85530	alcohol dehydrogen
25	27	75.0	373	1 A33419	alcohol dehydrogen
26	27	75.0	373	1 D67A	alcohol dehydrogen
27	27	75.0	373	1 S68061	alcohol dehydrogen
28	27	75.0	374	1 D67UC2	alcohol dehydrogen
29	27	75.0	374	1 A56643	alcohol dehydrogen

30	27	75.0	375	1 I55359	alcohol dehydrogen
31	27	75.0	375	1 D67UAB	alcohol dehydrogen
32	27	75.0	375	1 A33909	alcohol dehydrogen
33	27	75.0	376	1 JC4967	alcohol dehydrogen
34	27	75.0	376	1 S51187	alcohol dehydrogen
35	27	75.0	378	1 A49662	alcohol dehydrogen
36	27	75.0	378	1 H64052	alcohol dehydrogen
37	27	75.0	378	2 F81097	probable alcohol d
38	27	75.0	379	1 S51357	alcohol dehydrogen
39	27	75.0	379	1 S71244	alcohol dehydrogen
40	27	75.0	381	1 JN0447	alcohol dehydrogen
41	27	75.0	381	2 T03289	formaldehyde dehyd
42	27	75.0	381	2 T04164	formaldehyde dehyd
43	27	75.0	386	1 S31140	alcohol dehydrogen
44	27	75.0	396	1 S31959	acyl-lacyl-carrier
45	27	75.0	424	2 D75330	probable beta-lact
46	27	75.0	455	2 C84133	glutamine syntheta
47	27	75.0	475	2 T02313	endoplasmic reticu
48	27	75.0	475	2 F86415	hypothetical prote
49	27	75.0	475	2 B96816	hypothetical prote
50	27	75.0	488	2 B69415	group II decarboxy

ALIGNMENTS

RESULT 1
F83354
Probable sulfatase PA2333 [Imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F83354
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapid, K.; L.
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A62950; MUID:20437337
A:Accession: F83354
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-538 <STO>
A:Cross-references: GB:AE004659; GB:AE004091; NID:99948361; PIDN:AMG05721.1; GSPDB:GN
A:Experimental source: Strain PA01
C:Genetics:
A:Gene: PA2333

Query Match 83.3%; Score 30; DB 2; Length 538;
Best Local Similarity 62.5%; Pred. No. 36;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxyxg 8
Db 455 WKYIAYDG 462

RESULT 2
S20901
titin - rabbit (fragment)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 18-Jun-1999
C:Accession: S20901; I46520
R:Labelt, S.; Gautel, M.; Lakey, A.; Trinick, J.
EMBO J. 11, 1711-1716, 1992
A:Title: Towards a molecular understanding of titin.
A:Reference number: S20897; MUID:92258380
A:Accession: S20901
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-6805 <LAB>
A:Cross-references: EMBL:X64696
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1992

R:Labelt, S.; Barlow, D.P.; Gautel, M.; Gibson, T.; Holt, J.; Hsieh, C.L.; Francke, U.; Nature 345, 273-276, 1990
 A:Title: A regular pattern of two types of 100-residue motif in the sequence of titin.
 A:Reference number: 146520; MUID:90238553
 A:Accession: 146520
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 4235-5250 <LAB2>
 A:Cross-references: EMBL:X17329; NID:g1756; PIDN:CAA35207.1; PID:g930251
 C:Superfamily: titlin; fibronectin type III repeat homology; immunoglobulin homology; pro
 C:Keywords: muscle

Query Match 80.6%; Score 29; DB 2; Length 6805;
 Best Local Similarity 62.5%; Pred. No. 7.3e+02;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxxxxyxg 8
 || || |
 Db 6485 WKKPAYDG 6492

RESULT 3
 138344
 titin, cardiac muscle [validated] - human
 N:Alternate names: connectin
 N:Contains: serine/threonine-specific protein kinase (EC 2.7.1.-)
 C:Species: Homo sapiens (man)
 C>Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 15-Sep-2000
 C:Accession: 138344; 138345; S20898; S20897; S20899; S63665; S37393
 R:Labelt, S.; Kolmerer, B.
 Science 270, 293-296, 1995
 A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
 A:Reference number: A57430; MUID:96026330
 A:Accession: 138344
 A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/EM
 A:Molecule type: mRNA
 A:Residues: 1-26926 <LAB1>
 A:Cross-references: EMBL:X90568; NID:g1017424; PID:g1017425
 R:Musco, G.; Tlatziros, C.; Schuck, P.; Pastore, A.
 Biochemistry 34, 553-561, 1995
 A:Title: Dissecting titin into its structural motifs: identification of an alpha-helix m
 A:Reference number: 138345; MUID:95119041
 A:Accession: 138345
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1977-2014 <MUS>
 A:Cross-references: EMBL:X83270; NID:g602579; PIDN:CAA58243.1; PID:g602580
 A:Note: conformation and properties are reported for a synthetic peptide corresponding t
 R:Labelt, S.; Gautel, M.; Lakey, A.; Trinick, J.
 EMBO J. 11, 1711-1716, 1992
 A:Title: Towards a molecular understanding of titin.
 A:Reference number: S20897; MUID:92258380
 A:Accession: S20898
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 13597-14200, 'I', 14202-14696 <LAB2>
 A:Cross-references: EMBL:X64698; NID:g37192; PIDN:CAA45939.1; PID:g37193
 A:Accession: S20897
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: mRNA
 A:Residues: 16330-16382, 'S', 16384-16756, 'F', 16758-16860 <LAB3>
 A:Cross-references: EMBL:X64699; NID:g37190; PIDN:CAA45940.1; PID:g37191
 A:Accession: S20899
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: mRNA
 A:Residues: 'P', 22278-22431, 'R', 22433-22448, 'G', 22450-22453, 'Q', 22455-22480, 'TR', 22483-2
 A:Cross-references: EMBL:X64697; NID:g37190; PIDN:CAA45938.1; PID:g37195
 R:Kolmerer, B.; Olivieri, N.; Wilt, C.C.; Herrmann, B.G.; Labelt, S.
 J. Mol. Biol. 256, 556-563, 1996
 A:Title: Genomic organization of M line titin and its tissue-specific expression in two
 A:Reference number: S63665; MUID:96177761
 A:Accession: S63665

A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 26729-26825 <KOI>
 A:Cross-references: EMBL:X92412; NID:g1236761
 R:Gautel, M.; Leonard, K.; Labelt, S.
 EMBO J. 12, 3827-3834, 1993
 A:Title: Phosphorylation of KSP motifs in the C-terminal region of titin in different
 A:Reference number: S37393; MUID:94008990
 A:Accession: S37393
 A:Molecule type: mRNA
 A:Residues: 26831-26926 <GAU>
 R:Improta, S.; Politou, A.S.; Pastore, A.
 submitted to the Brookhaven Protein Data Bank, February 1996
 A:Reference number: A66736; PDB:1RTT
 A:Contents: annotation; conformation by (1)H-NMR, residues 5253-5341
 R:Pfuhl, M.; Pastore, A.
 submitted to the Brookhaven Protein Data Bank, August 1996
 A:Reference number: A66201; PDB:1NCT
 A:Contents: annotation; conformation by (1)H-NMR, residues 'S', 26059-26155
 C:Genetics:
 A:Gene: GDB:TTN
 A:Cross-references: GDB:127867; OMIM:188840
 A:Map position: 2q31-2q32
 C:Function:
 A:Description: structural protein forming filaments in striated muscle
 C:Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology;
 C:Keywords: alternative splicing; calmodulin binding; cardiac muscle; duplication; gl
 structural protein
 F:474752-25008/Domain: protein kinase homology <KIN>
 F:84,177,905,2276,2378,2459,2481,2563,2669,2763,2896,3088,3179,3384,3432,3628,3772,40
 98,11066,11488,11515,11635,11949,12170,12478,12526,12645,12875,13001,13036,13295,1354
 tatus Predicted
 F:16780,16976,17579,17602,17667,17681,17845,17899,18121,18188,18209,18336,18670,18680
 ,21900,21935,22255,22495,22627,22697,23024,23318,23883,24012,24177,24290,24447,24642,
 F:26171,26178,26184,26190/Binding site: Phosphate (Ser) (covalent) #status experiment

Query Match 80.6%; Score 29; DB 1; Length 26926;
 Best Local Similarity 62.5%; Pred. No. 2.8e+03;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxxxxyxg 8
 || || |
 Db 18126 WKKPAYDG 18133

RESULT 4
 S23164
 light-harvesting protein alpha chain - Ectothiorhodospira halochloris
 N:Alternate names: antenna pigment protein alpha chain
 C:Species: Ectothiorhodospira halochloris
 C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 21-Aug-1998
 C:Accession: S23164
 R:Wagner-Huber, R.; Brunisholz, R.A.; Blaisig, I.; Frank, G.; Suter, F.; Zuber, H.
 Eur. J. Biochem. 205, 917-925, 1992
 A:Title: The primary structure of the antenna polypeptides of Ectothiorhodospira halo
 A:Reference number: S23164; MUID:92249336
 A:Accession: S23164
 A:Molecule type: protein
 A:Residues: 1-65 <WAG>
 C:Superfamily: light-harvesting protein alpha chain
 C:Keywords: antenna complex; bacteriochlorophyll; blocked amino end; light-harvesting
 F:1/Modified site: N-formylmethionine #status experimental

Query Match 77.8%; Score 28; DB 2; Length 65;
 Best Local Similarity 50.0%; Pred. No. 13;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxxxxyxg 8
 || || |
 Db 58 WKKPAYDG 65

```
RESULT 5
A:Species: Brucella abortus
C:Species: Brucella abortus
C:Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 08-Oct-1999
A:Accession: A56152
R:de Wergifosse, P.; Lintermans, P.; Linet, J.N.; Cloeckaert, A.
J. Bacteriol. 177, 1911-1914, 1995
A:Title: Cloning and nucleotide sequence of the gene coding for the major 25-kilodalton
A:Reference number: A56152; MUID:95204367
A:Accession: A56152
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-213 <DEA>
A:Cross-references: GB:X79284; NID:g769744; PIDN:CAA55872.1; PID:g769745

Query Match      77.8%; Score 28; DB 2; Length 213;
Best Local Similarity 50.0%; Pred. No. 41;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxayxg 8
|||
Db 74 WKAGAFAG 81

RESULT 6
J01386
Hypothetical 33k protein - pea enation mosaic virus
C:Species: pea enation mosaic virus, PEMV
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 08-Oct-1999
A:Accession: J01386
R:Demler, S.A.; de Zoeten, G.A.
J. Gen. Virol. 72, 1819-1824, 1991
A:Title: The nucleotide sequence and luteovirus-like nature of RNA 1 of an aphid non-tre
A:Reference number: J01386; MUID:91341468
A:Accession: J01386
A:Molecule type: genomic RNA
A:Residues: 1-303 <DEM>
A:Cross-references: GB:I04573; NID:g294105; PIDN:AA72298.1; PID:g294107
A:Experimental source: strain MSG
A:Note: this reading frame extends between two stop codons and does not begin with a sta
A:Note: 65-Asp was also found
C:Genetics:
A:Map position: segment RNA1

Query Match      77.8%; Score 28; DB 2; Length 303;
Best Local Similarity 50.0%; Pred. No. 58;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxayxg 8
|||
Db 135 WQARVYTC 142

RESULT 7
T48858
Acetylpolymine aminohydrolase [validated] - Mycoplana ramosa
C:Species: Mycoplana ramosa
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
A:Accession: T48858
R:Sakurada, K.; Ohata, T.; Fujishiro, K.; Hasegawa, M.; Aisaka, K.
J. Bacteriol. 178, 5781-5786, 1996
A:Title: Acetylpolymine amidohydrolase from Mycoplana ramosa: gene cloning and characte
A:Reference number: Z24559
A:Accession: T48858
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-341 <SAK>
A:Cross-references: EMBL:D10463; PIDN:BA01256.1
A:Experimental source: ATCC 49678; strain FERM BP-1845
```

```
C:Genetics:
A:Note: aphA
C:Function:
A:Description: involved in degradation of acetylpolymines [validated; MUID:96422009]

Query Match      77.8%; Score 28; DB 2; Length 341;
Best Local Similarity 50.0%; Pred. No. 66;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wkxxayxg 8
|||
Db 78 WKAGGYKC 85

RESULT 8
D64884
Conserved hypothetical protein b1345 - Escherichia coli
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 29-Sep-1999
A:Accession: D64884
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: D64884
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-411 <BLAT>
A:Cross-references: GB:AE000232; GB:U00096; NID:g1787600; PIDN:AACT4427.1; PID:g17876
A:Experimental source: strain K-12, substrain MG1655
C:Superfamily: hypothetical protein b1579

Query Match      77.8%; Score 28; DB 2; Length 411;
Best Local Similarity 50.0%; Pred. No. 79;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wkxxayxg 8
|||
Db 222 WKFAVYSG 229

RESULT 9
H85748
Probable integrase for prophage CP-933R intR [imported] - Escherichia coli (strain O1
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
A:Accession: H85748
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: H85748
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-411 <STO>
A:Cross-references: GB:AE005174; NID:g12515407; PIDN:AAG56452.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: intR
C:Superfamily: hypothetical protein b1579

Query Match      77.8%; Score 28; DB 2; Length 411;
Best Local Similarity 50.0%; Pred. No. 79;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wkxxayxg 8
|||
Db 222 WKFAVYSG 229
```

```

RESULT 10
S60618
4-alpha-glucanotransferase (EC 2.4.1.25) - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 27-Apr-1996 #sequence_revision 08-Nov-1996 #text_change 21-Jul-2000
C:Accession: S60618; S23198; D72387
R:Heinrich, P.; Huber, W.; Liebl, W.
Syst. Appl. Microbiol. 17, 297-305, 1994
A:Title: Expression in Escherichia coli and structure of the gene encoding 4-alpha-glucose
related enzyme subfamilies.
A:Reference number: S60618
A:Accession: S60618
A:Molecule type: DNA
A:Residues: 1-441 <HEX>
A:Cross-references: EMBL:Z50813; NID:g951310; PIDN:CAA90693.1; PID:g951311
R:Liebl, W.; Fell, R.; Gabelsberger, J.; Kellermann, J.; Schleifer, K.H.
Eur. J. Biochem. 207, 81-88, 1992
A:Title: Purification and characterization of a novel thermostable 4-alpha-glucanotransf
A:Reference number: S23198; MUID:92331687
A:Accession: S23198
A:Molecule type: protein
A:Residues: 'A',2-23 <LIE>
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316
A:Accession: D72387
A:Molecule type: DNA
A:Residues: 1-264, 'G',266-441 <ARN>
A:Cross-references: GB:AF001716; GB:AE000512; NID:g4980853; PIDN:AD35451.1; PID:g498086
A:Experimental source: strain MSB8
A:Genetics: TM0364
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match      77.8%; Score 28; DB 2; Length 441;
Best Local Similarity 62.5%; Pred. No. 84;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY      1 wkxxayxg 8
      || || |
Db      348 WKMPAYNG 355

RESULT 11
A35742
aqualysin (EC 3.4.21.-) I precursor - Thermus aquaticus
C:Species: Thermus aquaticus
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: A35742; S00620; S00324
R:Terada, I.; Kwon, S.T.; Miyata, Y.; Matsuzawa, H.; Ohta, T.
J. Biol. Chem. 265, 6576-6581, 1990
A:Title: Unique precursor structure of an extracellular protease, aqualysin I, with NH-2
A:Reference number: A35742; MUID:90216674
A:Accession: A35742
A:Molecule type: DNA
A:Residues: 1-513 <TER>
A:Cross-references: GB:J90108; GB:D90108; GB:J05414; NID:q21717; PIDN:BA14135.1; PID:g
R:Kwon, S.T.; Terada, I.; Matsuzawa, H.; Ohta, T.
Eur. J. Biochem. 173, 491-497, 1988
A:Title: Nucleotide sequence of the gene for aqualysin I (a thermophilic alkaline serine
A:Reference number: S00620; MUID:88225062
A:Accession: S00620
A:Molecule type: DNA
A:Residues: 75-442 <KWO>
A:Cross-references: EMBL:X07734; NID:g48069; PIDN:CAA30559.1; PID:g602091
A:Note: part of this sequence, including the amino and carboxyl ends of the mature prote

```

```

R:Matsuzawa, H.; Tokugawa, K.; Hamaoki, M.; Mizoguchi, M.; Taguchi, H.; Terada, I.; K
Eur. J. Biochem. 171, 441-447, 1988
A:Title: Purification and characterization of aqualysin I (a thermophilic alkaline se
A:Reference number: S00324; MUID:88151937
A:Accession: S00324
A:Molecule type: protein
A:Residues: 128-170 <MATS>
C:Superfamily: subtilisin; subtilisin homology
C:Keywords: extracellular protein; hydrolase; serine proteinase
F:1-14/Domain: signal sequence #status predicted <SIG>
F:15-127/Domain: propeptide #status predicted <PRO>
F:128-408/Product: aqualysin I #status experimental <MAT>
F:157-364/Domain: subtilisin homology <SBI>
F:255-257,281-283/Region: SI specificity crevice #status predicted
F:409-513/Domain: carboxyl-terminal propeptide #status predicted <CPR>
F:166,197,349/Active site: Asp, His, Ser #status predicted

```

```

Query Match      77.8%; Score 28; DB 1; Length 513;
Best Local Similarity 50.0%; Pred. No. 98;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY      1 wkxxayxg 8
      || || |
Db      495 WRIVAYSG 502

RESULT 12
T31102
filamentous hemagglutinin 1 - Haemophilus ducreyi
C:Species: Haemophilus ducreyi
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T31102
R:Ward, C.K.; Lumbley, S.R.; Latimer, J.L.; Cope, L.D.; Hansen, E.J.
J. Bacteriol. 180, 6013-6022, 1998
A:Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.
A:Reference number: Z20984; MUID:99030326
A:Accession: T31102
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4152 <WAR>
A:Cross-references: EMBL:AF057695; NID:g3929017; PID:g3929018; PIDN:AAC79757.1
A:Genetics: lspa1
A:Gene: lspa1

```

```

Query Match      77.8%; Score 28; DB 2; Length 4152;
Best Local Similarity 50.0%; Pred. No. 7.6e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      1 wkxxayxg 8
      || || |
Db      2999 WKSGYAG 3006

```

```

RESULT 13
T31105
hypothetical protein 2 - Haemophilus ducreyi
C:Species: Haemophilus ducreyi
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T31105
R:Ward, C.K.; Lumbley, S.R.; Latimer, J.L.; Cope, L.D.; Hansen, E.J.
J. Bacteriol. 180, 6013-6022, 1998
A:Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.
A:Reference number: Z20984; MUID:99030326
A:Accession: T31105
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4919 <WAR>
A:Cross-references: EMBL:AF057696; NID:g3929021; PID:g3929023; PIDN:AAC79761.1
A:Genetics: lspa2
A:Gene: lspa2

```

Query Match 77.8%; Score 28; DB 2; Length 4919;
Best Local Similarity 50.0%; Pred. No. 9e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wxxxxxyxg 8
|| |
DB 3128 MKKSGVAG 3135

RESULT 14
D82576
hypothetical protein XF2277 [Imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: D82576
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: D82576
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-109 <SIM>
A:Cross-references: GB:AE004040; GB:AE003849; NID:g9107437; PIDN:AAF85076.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Stimpson, A.J.G.; Relnach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Brlones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carter, H
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
chudo, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, B
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J.; de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sava
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Teshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF2277

Query Match 75.0%; Score 27; DB 2; Length 109;
Best Local Similarity 50.0%; Pred. No. 36;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wxxxxxyxg 8
|| |
DB 25 WKLIGYSC 32

RESULT 15
QOCV2
aphid transmission protein - cauliflower mosaic virus
N:Alternate names: ORF II; ORF2 protein
C:Species: cauliflower mosaic virus
C:Date: 31-Oct-1980 #sequence_revision 31-Oct-1980 #text_change 23-Jul-1999
C:Accession: A90799; JAO010; A93729; A94613; JN0494; A04157
R:Frank, A.; Guillely, H.; Jonard, G.; Richards, K.; Hirth, L.
Cell 21, 285-294, 1980
A:Title: Nucleotide sequence of cauliflower mosaic virus DNA.
A:Reference number: A90799; MUID:81001865
A:Accession: A90799
A:Molecule type: DNA
A:Residues: 1-159 <FRA>
A:Cross-references: GB:W00141; GB:J02048; NID:g58821; PIDN:CAA23457.1; PID:g58823
A:Experimental source: strain Strasbourg
R:Modjtahedi, N.; Volovitch, M.; Mazzolini, L.; Yot, P.
FEBS Lett. 181, 223-228, 1985
A:Title: Comparison of the predicted secondary structure of aphid transmission factor fo

A:Reference number: A91334
A:Accession: JAO010
A:Molecule type: DNA
A:Residues: 1-159 <MOD>
A:Experimental source: strain PV14
R:Gardner, R.C.; Howarth, A.J.; Hahn, P.; Brown-Luedi, M.; Shepherd, R.J.; Messing, J
Nucleic Acids Res. 9, 2871-2888, 1981
A:Title: The complete nucleotide sequence of an infectious clone of cauliflower mosai
A:Reference number: A93729; MUID:82014878
A:Accession: A93729
A:Molecule type: DNA
A:Residues: 1-88, 'N', 90-93, 'R', 95-104, 'V', 106-117, 'N', 119-120, 'N', 123-126, 'D', 128-159
A:Experimental source: strain CM1841
R:Guillely, H.
submitted to the Nucleic Acid Sequence Database, October 1982
A:Reference number: A94613
A:Accession: A94613
A:Molecule type: DNA
A:Residues: 1-50, 'K', 52-101, 'P', 103-126, 'DE', 129-159 <GUT>
A:Experimental source: strain D/H
R:Chenault, K.D.; Melcher, U.
Gene 123, 255-257, 1993
A:Title: The complete nucleotide sequence of cauliflower mosaic virus isolate BBC.
A:Reference number: JN0493; MUID:93154593
A:Accession: JN0494
A:Molecule type: DNA
A:Residues: 1, 'R', 3-117, 'N', 119-126, 'D', 128-137, 'K', 139-159 <CHE>
A:Cross-references: GB:M00542; NID:g678542; PIDN:AAA02372.1; PID:g293182
A:Experimental source: isolate BBC
C:Superfamily: cauliflower mosaic virus aphid transmission protein

Query Match 75.0%; Score 27; DB 1; Length 159;
Best Local Similarity 50.0%; Pred. No. 52;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxyxg 8
|| |
DB 61 WKINSYFG 68

RESULT 16
T32510
hypothetical protein C44B12.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 04-Mar-2000
C:Accession: T32510
R:Tin-Mollam, A.
submitted to the EMBL Data Library, December 1997
A:Description: The sequence of C. elegans cosmid C44B12.
A:Reference number: Z21183
A:Accession: T32510
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-246 <TIN>
A:Cross-references: EMBL:AF036692; PIDN:AAB88326.1; GSPDB:GN000022; CESP:C44B12.3
C:Genetics:
A:Gene: CESP:C44B12.3
A:Map position: 4
A:Introns: 73/3, 112/3, 137/1, 202/1
C:Superfamily: Caenorhabditis elegans hypothetical protein C44B12.3

Query Match 75.0%; Score 27; DB 2; Length 246;
Best Local Similarity 50.0%; Pred. No. 80;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxyxg 8
|| |
DB 87 WKOKSYDG 94

RESULT 17
F82120
Zinc ABC transporter, permease protein VC2083 [imported] - *Vibrio cholerae* (strain N1696)
C:Species: *Vibrio cholerae*
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: F82120
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
charidson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F.
l.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
N:ature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
A:Reference number: A82035; MUID:20406833
A:Accession: F82120
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-260 <HEI>
A:Cross-references: GB:AE004282; GB:AE003852; NID:99656626; PIDN:AAF95229.1; GSPDB:GN001
C:Genetics:
A:Gene: VC2083
A:Map position: 1
C:Superfamily: conserved hypothetical protein HI0360

Query Match 75.0%; Score 27; DB 2; Length 260;
Best Local Similarity 50.0%; Pred. No. 85;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxayxg 8
|: || |
Db 28 WRRMAYFG 35

RESULT 18
A64066
Probable membrane protein HI0407 - *Haemophilus influenzae* (strain Rd KW20)
C:Species: *Haemophilus influenzae*
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C:Accession: A64066
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
charidson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F.
l.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
N:ature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
A:Reference number: A82035; MUID:20406833
A:Accession: F82120
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-260 <HEI>
A:Cross-references: GB:AE004282; GB:AE003852; NID:99656626; PIDN:AAF95229.1; GSPDB:GN001
C:Genetics:
A:Gene: VC2083
A:Map position: 1
C:Superfamily: conserved hypothetical protein HI0360

Query Match 75.0%; Score 27; DB 2; Length 261;
Best Local Similarity 50.0%; Pred. No. 85;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxayxg 8
|: || |
Db 28 WRRMAYFG 35

RESULT 19
F82959
Permease of ABC zinc transporter ZnuB PA5501 [imported] - *Pseudomonas aeruginosa* (str
C:Species: *Pseudomonas aeruginosa*
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F82959
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
N:ature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337
A:Accession: F82959
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-262 <STO>
A:Cross-references: GB:AE004962; GB:AE004091; NID:99951826; PIDN:AA08886.1; GSPDB:GN
C:Genetics:
A:Gene: znuB; PA5501
C:Superfamily: conserved hypothetical protein HI0360

Query Match 75.0%; Score 27; DB 2; Length 262;
Best Local Similarity 50.0%; Pred. No. 85;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxayxg 8
|: || |
Db 28 WRRMAYFG 35

RESULT 20
D75286
Serine proteinase truncated homolog DR2322 [imported] - *Deinococcus radiodurans* (stra
C:Species: *Deinococcus radiodurans*
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Sep-2000
C:Accession: D75286
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
N:ature 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* RI.
A:Reference number: A75250; MUID:20036896
A:Accession: D75286
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-288 <WHI>
A:Cross-references: GB:AE002064; GB:AE000513; NID:96460134; PIDN:AAF11870.1; PID:9646
A:Experimental source: strain RI
C:Genetics:
A:Gene: DR2322
A:Map position: 1

Query Match 75.0%; Score 27; DB 2; Length 288;
Best Local Similarity 50.0%; Pred. No. 94;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxayxg 8
|: || |
Db 271 WEVYATSG 278

Search completed: January 14, 2002, 07:37:26
Job time: 896 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:40:37 ; Search time 37.71 Seconds
(without alignments)
7.778 Million cell updates/sec

Title: 09-185908-1B
Perfect score: 36
Sequence: 1 wkxxxyxg 8

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB ID	Description
1	31	86.1	215	1	CLDX_BRARE
2	31	86.1	1018	1	HTRA_CHITK
3	29	80.6	230	1	CLD2_HUMAN
4	28	77.8	65	1	LHA2_ECTHL
5	28	77.8	201	1	OM25_BRUVY
6	28	77.8	211	1	CLD1_MOUSE
7	28	77.8	211	1	CLD1_RAT
8	28	77.8	213	1	OM25_BRUBA
9	28	77.8	213	1	OM25_BRUCA
10	28	77.8	213	1	OM25_BRUNE
11	28	77.8	213	1	OM25_BRUNE
12	28	77.8	213	1	OM25_BRUNE
13	28	77.8	341	1	APHA_MYCRA
14	28	77.8	375	1	ADH1_GADMO
15	28	77.8	411	1	INTR_ECOLI
16	28	77.8	441	1	MCTA_THENA
17	28	77.8	513	1	AQOL1_THEAO
18	28	77.8	934	1	HMDH_PICJA
19	27	75.0	99	1	VAT_CAMVY
20	27	75.0	159	1	VAT_CAMVY
21	27	75.0	159	1	VAT_CAMVY
22	27	75.0	159	1	VAT_CAMVE
23	27	75.0	159	1	VAT_CAMVY
24	27	75.0	159	1	VAT_CAMVY
25	27	75.0	208	1	CLDX_BRARE
26	27	75.0	217	1	CLD9_HUMAN
27	27	75.0	217	1	CLD9_MOUSE
28	27	75.0	219	1	CLD6_MOUSE
29	27	75.0	220	1	CLD6_HUMAN
30	27	75.0	261	1	ZNUB_HAETN
31	27	75.0	369	1	ADH3_ECOLI
32	27	75.0	369	1	ADH3_PASPI
33	27	75.0	373	1	ADH3_HORSE

ALIGNMENTS

RESULT 1	CLDX_BRARE	STANDARD:	PRT:	215 AA.
ID	CLDX_BRARE			
AC	Q9YH92;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	CLAUDIN-LIKE PROTEIN ZFAA22 (CLAUSIN 7).			
OS	Brachydanio rerio (zebrafish) (Zebra danio).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;			
OC	Cypriniformes; Cyprinidae; Rasbora; Danio.			
OX	NCBI_TaxID=7955;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Keen T.J., Ingelhearn C.F.;			
RL	Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20525589; PubMed=11071763;			
RA	Chin A.-J., Tsang M., Weinberg E.S.;			
RT	"Heart and gut chiralities are controlled independently from initial			
RT	heart position in the developing zebrafish.";			
RL	Dev. Biol. 227:403-421(2000).			
CC	- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.			
CC	- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.			
CC	- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.			
CC	- This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
DR	EMBL: AF011788; GAA09776.1; -			
DR	EMBL: AF260240; AAG24512.1; -			
DR	InterPro: IPR001832; Claudin.			
DR	InterPro: IPR000729; PMP22_Claudin.			
DR	Pfam: PF008822; PMP22_Claudin. 1.			
DR	PRINTS: PRO1077; CLAUDIN.			
DR	PROSITE: PS01346; CLAUDIN. 1.			
KW	Tight junction; Transmembrane.			
FT	TRANSMEM 8			
FT	TRANSMEM 82			
FT	TRANSMEM 118			
FT	TRANSMEM 163			
FT	TRANSMEM 183			
SQ	SEQUENCE 215 AA: 22865 MW: BC04870B75B8CB9D CRC64;			

34	27	75.0	373	1	ADHX_HUMAN	P11766	homo sapien
35	27	75.0	373	1	ADHX_MOUSE	P28474	mus musculus
36	27	75.0	373	1	ADHX_RABIT	O19053	oryctolagus
37	27	75.0	373	1	ADHX_RAT	P12711	rattus norv
38	27	75.0	373	1	ADHX_UROSA	P80467	uromastix h
39	27	75.0	374	1	ADHB_HUMAN	P00325	homo sapien
40	27	75.0	374	1	ADH_MACMU	P28469	macaca mula
41	27	75.0	374	1	ADH_PAPHA	P14139	papio hamad
42	27	75.0	375	1	ADHL_GADMO	P81601	gadus morhu
43	27	75.0	375	1	FADH_PARDE	P43882	paracoccus
44	27	75.0	376	1	ADH1_RHOSH	P72324	rhodobacter
45	27	75.0	376	1	ADHX_MYXGL	P80360	myxine glut
46	27	75.0	376	1	ADHX_SPAU	P79886	sparus aura
47	27	75.0	378	1	ADH3_HAETN	P44357	haemophilus
48	27	75.0	378	1	ADHX_DROME	P46415	drosophila
49	27	75.0	378	1	ADHX_OCTVU	P81431	octopus vul
50	27	75.0	378	1	ADHX_PEA	P80572	pisum sativ

Query Match 86.1%; Score 31; DB 1; Length 215;
 Best Local Similarity 62.5%; Pred. No. 4.3;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxyg 8
 || || |
 Db 30 WKMSAYVG 37

RESULT 2
 HIRA_CHICK STANDARD; PRT; 1018 AA.
 ID HIRA_CHICK P79987;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 38, Last annotation update)
 DE HIRA PROTEIN (TUP1 LIKE ENHANCER OF SPLIT PROTEIN 1) (HIRA).
 GN HIRA OR TUPLE1.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97217783; PubMed=9063744;
 RA Roberts C., Daw S.C., Halford S., Scambler P.J.;
 RT "Cloning and developmental expression analysis of chick Hira (Chira),
 a candidate gene for Disgeorge syndrome.";
 RL Hum. Mol. Genet. 6:237-245(1997).
 CC -1- FUNCTION: COULD PLAY A PART IN MECHANISMS OF TRANSCRIPTIONAL
 CC REGULATION SIMILAR TO THAT PLAYED BY YEASTS HIR1 AND HIR2 TOGETHER.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: CONTAINS 4 WD REPEATS (TRP-ASP DOMAINS).
 CC -1- SIMILARITY: BELONGS TO THE HIR1 FAMILY OF WD-REPEAT PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X99375; CAA67754.1; -;
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40.7.
 DR PRINTS: PR00320; GPROTEINBRPT.
 DR SMART: SM00320; WD40; 6.
 DR PROSITE: PS00678; WD_REPEATS_1; 1.
 DR PROSITE: PS50082; WD_REPEATS_2; 3.
 DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
 KW Transcription regulation; Repeat; WD repeat; Nuclear protein.
 FT REPEAT 68 98 WD 1.
 FT REPEAT 128 158 WD 2.
 FT REPEAT 172 202 WD 3.
 FT REPEAT 266 313 WD 4.
 FT DOMAIN 267 286 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 634 651 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 407 413 POLY-GLN.
 FT DOMAIN 552 555 POLY-SER.
 FT DOMAIN 644 647 POLY-LYS.
 SO SEQUENCE 1018 AA; 111817 MW; 58C8C710A8FA7BF1 CRC64;

Query Match 86.1%; Score 31; DB 1; Length 1018;
 Best Local Similarity 62.5%; Pred. No. 19;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxyg 8
 || || |
 Db 97 WKRAAYIG 104

RESULT 3
 CLD2_HUMAN STANDARD; PRT; 230 AA.
 ID CLD2_HUMAN P57739;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE CLAUDIN-2.
 GN CLDN2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TRISUE-Epithelium;
 RA Reinecker H.-C.; Sakaguchi T.; Golden H.M.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,
 RA Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,
 RA Yu J., Han L.H.;
 RT "Novel human cDNA clone with function of inhibiting cancer cell
 RT growth.";
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF250558; AAF98151.1; -;
 DR EMBL: AF177340; AAG17984.1; -;
 DR InterPro: IPR001832; Claudin.
 DR Pfam: PF00822; PMP22_Claudin.
 DR PROSITE: PS01346; CLAUDIN; 1.
 KW Tight junction; Transmembrane.
 FT TRANSMEM 8 28 POTENTIAL.
 FT TRANSMEM 82 102 POTENTIAL.
 FT TRANSMEM 117 137 POTENTIAL.
 FT TRANSMEM 163 183 POTENTIAL.
 SO SEQUENCE 230 AA; 24548 MW; 52CA642D4A62B70D CRC64;

Query Match 80.6%; Score 29; DB 1; Length 230;
 Best Local Similarity 50.0%; Pred. No. 13;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxyg 8
 || || |
 Db 30 WKTSYVG 37

RESULT 4
 LHA2_ECTHL STANDARD; PRT; 65 AA.
 ID LHA2_ECTHL P80103;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE LIGHT-HARVESTING PROTEIN B800/830/1020, ALPHA-2 CHAIN (EHS-ALPHA-2)
 DE (ANTENNA PIGMENT PROTEIN, ALPHA-2 CHAIN).
 OS Ectothiorhodospira halochloris.
 OC Bacteria; Proteobacteria; gamma subdivision; Ectothiorhodospiraceae;

```
OC Haldichodospira.
OX NCBI_TaxID=1052;
RN [1]
RP SEQUENCE.
RC STRAIN=DSM 1059;
RX MEDLINE=92249336; PubMed=1577009;
RA Wagner-Huber R., Brunisholz R.A., Bissig I., Frank G., Suter F.,
  Zuber H.;
RT "The primary structure of the antenna polypeptides of
  ectochlorodospira halochloris and ectochlorodospira halophila. Four
  core-type antenna polypeptides in E. halochloris and E. halophila.";
RL Eur. J. Biochem. 205:917-925(1992).
CC -1- FUNCTION: ANTENNA COMPLEXES ARE LIGHT-HARVESTING SYSTEMS, WHICH
  TRANSFER THE EXCITATION ENERGY TO THE REACTION CENTERS.
CC -1- SUBUNIT: THE CORE COMPLEX IS FORMED BY DIFFERENT ALPHA AND BETA
  CHAINS, BINDING BACTERIOCHLOROPHYLL MOLECULES, AND ARRANGED
  MOST PROBABLY IN TETRAMERIC STRUCTURES DISPOSED AROUND THE
  REACTION CENTER. THE NON-PIGMENTED GAMMA CHAINS MAY CONSTITUTE
  ADDITIONAL COMPONENTS.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. INNER MEMBRANE.
CC PIR: S23164; S23164.
DR InterPro: IPR002361; Antenna_comp_alpha.
DR InterPro: IPR000066; LHC.
DR Pfam: PF00556; LHC; 1.
DR PROSITE: PS00968; ANTENNA_COMP_ALPHA; 1.
KW Antenna complex; Light-harvesting polypeptide; Transmembrane;
  Magnesium; Bacteriochlorophyll; Inner membrane.
FT DOMAIN 1 13 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 14 34 PERIPLASMIC (POTENTIAL).
FT DOMAIN 35 65 AXIAL LIGAND TO THE BACTERIOCHLOROPHYLL
FT METAL 29 MAGNESIUM (POTENTIAL).
SQ SEQUENCE 65 AA; 7688 MW; 55AAC30674BE3D9A CRC64;

Query Match 77.8%; Score 28; DB 1; Length 65;
Best Local Similarity 50.0%; Pred. No. 6.5;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 WKXXYXG 8
   |||
Db 58 WKRTSYDC 65

RESULT 5
OM25_BRUOV STANDARD; PRT; 201 AA.
AC Q45335;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 25 KDA OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
GN OMP25.
OS Brucella ovis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=236;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=63/290;
RC MEDLINE=96239016; PubMed=8675306;
RA Cloeckaert A., Verger J.M., Grayon M., Zygmunt M.S., Grepinet O.;
  "Nucleotide sequence and expression of the gene encoding the major
  25-kilodalton outer membrane protein of Brucella ovis: Evidence for
  antigenic shift, compared with other Brucella species, due to a
  deletion in the gene.";
RL Infect. Immun. 64:2047-2055(1996).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE OMP25 / ROB FAMILY. HAS A C-TERMINAL
  DELETION COMPARED TO THAT OF OTHER BRUCELLA SPECIES.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sb-sib.ch).
CC -----
DR EMBL: U33004; AB06702.1; -.
DR InterPro: IPR000498; OmpA_tmem.
DR Pfam: PF01389; OmpA_membrane; 1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 201
FT SEQUENCE 201 AA; 21817 MW; 85A4897489A0935B CRC64;

Query Match 77.8%; Score 28; DB 1; Length 201;
Best Local Similarity 50.0%; Pred. No. 19;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 WKXXYXG 8
   |||
Db 74 WKAGAFAG 81

RESULT 6
CLD1_MOUSE STANDARD; PRT; 211 AA.
ID CLD1_MOUSE
AC O88551;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-1.
GN CLDN1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98311639; PubMed=9647647;
RA Furuse M., Fujita K., Hiragi T., Fujimoto K., Tsukita S.;
  "Claudin-1 and -2: novel integral membrane proteins localizing at
  tight junctions with no sequence similarity to occludin.";
RL J. Cell Biol. 141:1539-1550(1998).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sb-sib.ch).
CC -----
DR EMBL: AF072127; AAC27078.1; -.
DR MGD: MGI:1276109; Cldn1.
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin; 1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN; 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 116 136 POTENTIAL.
FT TRANSMEM 164 184 POTENTIAL.
SQ SEQUENCE 211 AA; 22881 MW; BEF896FA62DBB6F0 CRC64;

Query Match 77.8%; Score 28; DB 1; Length 211;
```

```
Best Local Similarity 50.0%; Pred. No. 20;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxayxg 8
   || :| |
Db 30 WKIYSYAG 37

RESULT 7
CLD1_RAT
ID CLD1_RAT STANDARD; PRT; 211 AA.
AC P56745;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE CLAUDIN-1.
GN CLDN1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RA Gregory M., Dufresne J., Cyr D.G.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AF195500; AAF04850.1; -.
DR InterPro; IPR001832; Claudin.
DR InterPro; IPR000732; PMP22.Claudin.
DR Pfam; PF008822; PMP22.Claudin; 1.
DR PRINTS; PRO1077; CLAUDIN.
DR PROSITE; PS01346; CLAUDIN; 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 116 136 POTENTIAL.
FT TRANSMEM 164 184 POTENTIAL.
SQ SEQUENCE 211 AA; 22871 MW; 3595931636631DDC CRC64;

Query Match 77.8%; Score 28; DB 1; Length 211;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxayxg 8
   || :| |
Db 30 WKIYSYAG 37

RESULT 8
OM25_BRUAB
ID OM25_BRUAB STANDARD; PRT; 213 AA.
AC 044664;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 25 KDA OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
GN OMP25
OS Brucella abortus.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
```

```
OX NCBI_TaxID=235;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=544S;
RA MEDLINE=95204367; PubMed=7896724;
RA de Wergifosse P., Lintermans P., Limet J.N., Cloeckaert A.;
RT Cloning and nucleotide sequence of the gene coding for the major 25-
RT kDa outer membrane protein of Brucella abortus.";
RL J. Bacteriol. 177:1911-1914(1995)
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X79284; CA55872.1; -.
DR InterPro; IPR000498; OmpA_tmem.
DR Pfam; PF01389; OmpA_membrane; 1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 213 25 KDA OUTER-MEMBRANE IMMUNOGENIC
FT PROTEIN
SQ SEQUENCE 213 AA; 23052 MW; 2328515F1F794BC7 CRC64;

Query Match 77.8%; Score 28; DB 1; Length 213;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxayxg 8
   || :| |
Db 74 WKAGAFAG 81

RESULT 9
OM25_BRUCA
ID OM25_BRUCA STANDARD; PRT; 213 AA.
AC 045110;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 25 KDA OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
GN OMP25.
OS Brucella canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=36855;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RM6/66;
RA MEDLINE=96239016; PubMed=8675306;
RA Cloeckaert A., Verger J.M., Grayson M., Zygmunt M.S., Grepinet O.;
RT Nucleotide sequence and expression of the gene encoding the major
RT 25-kilodalton outer membrane protein of Brucella ovis: Evidence for
RT antigenic shift, compared with other Brucella species, due to a
RT deletion in the gene.";
RL Infect. Immun. 64:2047-2055(1996).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
```

```
DR EMBL: U39358; AAB36692.1; -
DR InterPro: IPR000498; OmpA_tmem.
DR Pfam: PF01389; OmpA_membrane; 1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 23
FT CHAIN 24 213 25 KDA OUTER-MEMBRANE IMMUNOGENIC
FT PROTEIN.
SQ SEQUENCE 213 AA: 23123 MW: D7E40E247A39B9DF CRC64:

Query Match 77.8%; Score 28; DB 1; Length 213;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxyxyxg 8
   |||
   74 WKAGAFAG 81

RESULT 10
OM25_BRUNE STANDARD; PRT; 213 AA.
AC Q45321:
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 25 KDA OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
OS Omp25.
GN Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID:29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:16M;
RX MEDLINE:96239016; PubMed:8675306;
RA Cloeckaert A., Verger J.M., Grayon M., Zygmunt M.S., Grepinet O.;
RT "Nucleotide sequence and expression of the gene encoding the major
RT 25-kilodalton outer membrane protein of Brucella ovis: Evidence for
RT antigenic shift, compared with other Brucella species, due to a
RT deletion in the gene."
RL Infect. Immun. 64:2047-2055(1996).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U33003; AAB06701.1; -
DR InterPro: IPR000498; OmpA_tmem.
DR Pfam: PF01389; OmpA_membrane; 1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 23
FT CHAIN 24 213 25 KDA OUTER-MEMBRANE IMMUNOGENIC
FT PROTEIN.
SQ SEQUENCE 213 AA: 23185 MW: 59003CF46F60DED3 CRC64:
```

```
OM25_BRUNE STANDARD; PRT; 213 AA.
ID OM25_BRUNE
AC Q45326;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 25 KDA OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
DE Omp25.
GN Brucella neotomae.
OS Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID:29460;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:5K33;
RX MEDLINE:96239016; PubMed:8675306;
RA Cloeckaert A., Verger J.M., Grayon M., Zygmunt M.S., Grepinet O.;
RT "Nucleotide sequence and expression of the gene encoding the major
RT 25-kilodalton outer membrane protein of Brucella ovis: Evidence for
RT antigenic shift, compared with other Brucella species, due to a
RT deletion in the gene."
RL Infect. Immun. 64:2047-2055(1996).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U39359; AAB36694.1; -
DR InterPro: IPR000498; OmpA_tmem.
DR Pfam: PF01389; OmpA_membrane; 1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 23
FT CHAIN 24 213 25 KDA OUTER-MEMBRANE IMMUNOGENIC
FT PROTEIN.
SQ SEQUENCE 213 AA: 23167 MW: 49EE0F47B784F87 CRC64:

Query Match 77.8%; Score 28; DB 1; Length 213;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxyxyxg 8
   |||
   74 WKAGAFAG 81

RESULT 12
OM25_BRUNE STANDARD; PRT; 213 AA.
AC Q45689;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 25 KDA OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
DE Omp25.
GN Brucella suis.
OS Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID:29461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:1330;
RX MEDLINE:96239016; PubMed:8675306;
RA Cloeckaert A., Verger J.M., Grayon M., Zygmunt M.S., Grepinet O.;
RT "Nucleotide sequence and expression of the gene encoding the major
RT 25-kilodalton outer membrane protein of Brucella ovis: Evidence for
RT antigenic shift, compared with other Brucella species, due to a
```

```

RT deletion in the gene."
RL Infect. Immun. 64:2047-2055(1996).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: U39397; AAB3695.1; -.
DR InterPro: IPR000498; OmpA.tmem.
DR Pfam: PF01389; OmpA_membrane.1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 213 25 KDA OUTER-MEMBRANE IMMUNOGENIC
FT PROTEIN.
SQ SEQUENCE 213 AA; 23151 MW; 55F33CF46FD0ED3 CRC64;

Query Match 77.8%; Score 28; DB 1; Length 213;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxyxg 8
| | | |
Db 74 WKAGAFAG 81

RESULT 13
APHA_MYCRA STANDARD; PRT; 341 AA.
ID APHA_MYCRA
AC Q48935;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ACETYLPOLYAMINE AMINOHYDROLASE.
GN APHA OR APH.
OS Mycoplasma ramosa (Mycoplasma bullata).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Mycoplasma.
OX NCBI_TaxID=40837;
[1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-30.
RC STRAIN-ATCC 49678 / NCIB 9440 / FERM BP-1845;
RX MEDLINE=96422009; PubMed=8824626;
RA Sakurada K., Ohta T., Fujishiro K., Hasegawa M., Aisaka K.;
RT "Acetylpolymine amidohydrolase from Mycoplasma ramosa: gene cloning
RT and characterization of the metal-substituted enzyme.";
RL J. Bacteriol. 178:5781-5786(1996).
[2]
RN CHARACTERIZATION.
RP STRAIN-ATCC 49678 / NCIB 9440 / FERM BP-1845;
RX MEDLINE=89087434; PubMed=3207420;
RA Fujishiro K., Ando M., Uwajima T.;
RT "Crystallization and some properties of acetylpolymine
RT amidohydrolase from Mycoplasma bullata.";
RL Biochem. Biophys. Res. Commun. 157:1169-1174(1988).
CC -1- FUNCTION: ACTS ON MANY TYPES OF ACETYLPOLYAMINES HAS HIGH AFFINITY
CC TOWARDS ACETYLPUTRESCINE, ACETYLCADAVERINE, ACETYLSPERMIDINE, AND
CC ACETYLSPERMINE. OPTIMAL PH AROUND 8.0.
CC -1- COFACTOR: BINDS ONE ZINC ION PER MOLECULE.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / APHA
CC FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: D10463; BAA01256.1; -.
DR InterPro: IPR000286; His_deacetylase.
DR Pfam: PF00850; Hist_deacetyl1.
DR PRINTS: PR01270; HDASUPER.
KW Hydrolase; Zinc.
SQ SEQUENCE 341 AA; 36332 MW; 950583DF79059F4A CRC64;

Query Match 77.8%; Score 28; DB 1; Length 341;
Best Local Similarity 50.0%; Pred. No. 32;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wxxyxg 8
| | | |
Db 78 WKAGYKG 85

RESULT 14
ADHH_GADMO STANDARD; PRT; 375 AA.
ID ADHH_GADMO
AC F81600;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ALCOHOL DEHYDROGENASE CLASS III H CHAIN (EC 1.1.1.1) (GLUTATHIONE-
DE DEPENDENT FORMALDEHYDE DEHYDROGENASE) (EC 1.2.1.1) (FDH).
OS Gadus morhua (Atlantic cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae;
OC Gadus.
OX NCBI_TaxID=8049;
[1]
RN SEQUENCE.
RX MEDLINE=97085413; PubMed=8931553;
RA Danielsson O., Shafigat J., Estonius M., El-Ahmad M., Joernvall H.;
RT "Isozyme multiplicity with anomalous dimer patterns in a class III
RT alcohol dehydrogenase. Effects on the activity and quaternary
RT structure of residue exchanges at 'non-functional' sites in a native
RT protein.";
RL Biochemistry 35:14561-14568(1996).
CC -1- FUNCTION: CLASS-III ADH IS REMARKABLY INEFFECTIVE IN OXIDIZING
CC ETHANOL, BUT IT READILY CATALYZES THE OXIDATION OF LONG-CHAIN
CC PRIMARY ALCOHOLS AND THE OXIDATION OF S-(HYDROXYMETHYL)
CC GLUTATHIONE.
CC -1- CATALYTIC ACTIVITY: ALCOHOL + NAD(+) = ALDEHYDE OR KETONE + NADH.
CC -1- CATALYTIC ACTIVITY: FORMALDEHYDE + GLUTATHIONE + NAD(+) =
CC S-FORMYLGLUTATHIONE + NADH.
CC -1- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY.
CC -1- SUBUNIT: HOMODIMER AND HETERODIMER WITH L CHAIN.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY. CLASS-III SUBFAMILY.
DR InterPro: IPR002328; ADH_zinc.
DR InterPro: IPR002085; ADH_zinc.
DR Pfam: PF00107; adh_zinc.1.
DR PROSITE: PS00059; ADH_ZINC; 1.
KW Oxidoreductase; Zinc; NAD; Multigene family; Acetylation.
FT MOD_RES 1 1 ACETYLATION.
FT METAL 46 46 ZINC (CATALYTIC).
FT METAL 68 68 ZINC (CATALYTIC).
FT METAL 98 98 ZINC (SECOND ATOM).
FT METAL 101 101 ZINC (SECOND ATOM).
FT METAL 104 104 ZINC (SECOND ATOM).
FT METAL 112 112 ZINC (SECOND ATOM).
FT METAL 115 115 ZINC (CATALYTIC).
FT BINDING 116 116 IMPORTANT FOR FDH ACTIVITY AND ACTIVATION
FT BY FATTY ACIDS (BY SIMILARITY).
SQ SEQUENCE 375 AA; 39669 MW; 0B9760AB77329FE3 CRC64;

```

Query Match 77.8%; Score 28; DB 1; Length 375;
 Best Local Similarity 50.0%; Pred. No. 35;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxyxg 8
 DB 315 WKATAFGC 322

RESULT 15
 INTR_ECOLI STANDARD; PRT; 411 AA.
 ID INTR_ECOLI
 AC P76056;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PUTATIVE LAMBDOID PROPHAGE RAC INTEGRASE.
 GN INTR OR B1345.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RA "The complete genome sequence of Escherichia coli K-12";
 RL Science 277:1453-1474(1997).
 RT -1- FUNCTION: INTEGRASE IS NECESSARY FOR INTEGRATION OF THE PHAGE INTO
 CC THE HOST GENOME BY SITE-SPECIFIC RECOMBINATION. IN CONJUNCTION
 CC WITH EXCISIONASE, INTEGRASE IS ALSO NECESSARY FOR EXCISION OF THE
 CC PROPHAGE FROM THE HOST GENOME (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE "PHAGE" INTEGRASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; AE000232; AAC74427.1; -.
 DR ECGene; EG13358; Intr.
 DR InterPro: IPR002104; Phage_Integrase.
 KW Pfam: PF00589; Phage_Integrase; 1.
 KW Hypothetical protein; DNA recombination; DNA integration;
 KW Complete Proteome.
 SQ SEQUENCE 411 AA; 47539 MW; E14B91630E0EF9A5 CRC64;

Query Match 77.8%; Score 28; DB 1; Length 411;
 Best Local Similarity 50.0%; Pred. No. 38;
 Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wxxxxxyxg 8
 DB 222 WKPAVYSG 229

RESULT 16
 MGTA_THEME STANDARD; PRT; 441 AA.
 ID MGTA_THEME
 AC P80099;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE 4-ALPHA-GLUCANOTRANSFERASE (EC 2.4.1.25) (AMYLALYASE)
 DE (DISPROPORTIONATING ENZYME) (D-ENZYME).
 GN MGTA OR TM0364.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogales; Thermotoga.
 OX NCBI_TaxID=2336;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=MSB8 / DSM 3109;
 RX MEDLINE=99287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
 RA Halt D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback T.R., Malek J.A., Linner K.D., Garrett M.K.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from
 RT genome sequence of Thermotoga maritima";
 RL Nature 399:323-329(1999).
 RN (12)
 RP SEQUENCE OF 1-23, AND CHARACTERIZATION.
 RC STRAIN=MSB8 / DSM 3109;
 RX MEDLINE=92331687; PubMed=1628664;
 RA Liebel W., Fell R., Gabelsberger J., Kellermann J., Schaefer K.-H.;
 RT "Purification and characterization of a novel thermostable 4-alpha-
 RT glucanotransferase of Thermotoga maritima cloned in Escherichia
 RT coli";
 RL Eur. J. Biochem. 207:81-88(1992).
 CC -1- CATALYTIC ACTIVITY: TRANSFERS A SEGMENT OF A 1,4-ALPHA-D-GLUCAN
 CC TO A NEW 4-POSITION IN AN ACCEPTOR, WHICH MAY BE GLUCOSE OR
 CC 1,4-ALPHA-D-GLUCAN.
 CC -1- SUBUNIT: MONOMER.
 CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
 CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; AE001716; AAD35451.1; -.
 DR PIR: S23198; S23198.
 DR TIGR: TM0364; -.
 DR InterPro: IPR000461; Alpha_amylase.
 DR Pfam: PF00128; alpha-amylase; 1.
 DR PRINTS: PR00110; ALPHAAMYLASE.
 KW Transferase; Glycosyltransferase; Carbohydrate metabolism;
 KW Complete proteome.
 FT ACT_SITE 216 BY SIMILARITY.
 FT ACT_SITE 278 BY SIMILARITY.
 FT CONFLECT 1 M -> A (IN REF. 2).
 SQ SEQUENCE 441 AA; 51843 MW; 9FB4C2AB09D3DF3A CRC64;

Query Match 77.8%; Score 28; DB 1; Length 441;
 Best Local Similarity 62.5%; Pred. No. 41;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxyxg 8
 DB 348 WKPAVYNG 355

RESULT 17
 AQL1_THEME STANDARD; PRT; 513 AA.
 ID AQL1_THEME
 AC P08594;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)

DT 01-NOV-1995 (Rel. 32, last annotation update)
 DE AQUALYSIN I PRECURSOR (EC 3.4.21.-).
 GN PST.
 OS Thermus aquaticus.
 OC Bacteria; Thermus/Delnooccus group; Thermus group; Thermus.
 OX NCBI_TaxId=271;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 15-23.
 RC STRAIN-YT1:
 RX MEDLINE=90216674; PubMed=2182621;
 RA Terada I., Kwon S.-T., Miyata Y., Matsuzawa H., Ohta T.;
 RT "Unique precursor structure of an extracellular protease, aqualysin
 I, with NH2- and COOH-terminal pro-sequences and its processing in
 Escherichia coli.";
 RL J. Biol. Chem. 265:6576-6581(1990).
 RN [2]
 RP SEQUENCE OF 75-442 FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-YT1:
 RX MEDLINE=88225062; PubMed=3286255;
 RA Kwon S.-T., Terada I., Matsuzawa H., Ohta T.;
 RT "Nucleotide sequence of the gene for aqualysin I (a thermophilic
 alkaline serine protease) of Thermus aquaticus YT-1 and
 characteristics of the deduced primary structure of the enzyme.";
 RL Eur. J. Biochem. 173:491-497(1988).
 RN [3]
 RP SEQUENCE OF 128-170.
 RX MEDLINE=88151937; PubMed=3162211;
 RA Matsuzawa H., Tokugawa K., Hamaki M., Mizoguchi M., Taguchi H.,
 Terada I., Kwon S.-T., Ohta T.;
 RT "Purification and characterization of aqualysin I (a thermophilic
 alkaline serine protease) produced by Thermus aquaticus YT-1.";
 RL Eur. J. Biochem. 171:441-447(1988).
 CC -1- FUNCTION: AQUALYSIN I IS A THERMOPHILIC ALKALINE SERINE PROTEASE.
 THE OPTIMAL TEMPERATURE FOR ITS CASEINOLYTIC ACTIVITY IS 80
 DEGREES CELSIUS.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- DEVELOPMENTAL STAGE: SECRETED FROM THE EARLY STATIONARY PHASE
 UNTIL THE TIME THE CELLS CEASE TO GROW.
 CC -1- PM: THE N- AND C-TERMINAL PRO-SEQUENCES ARE REMOVED THROUGH THE
 PROTEOLYTIC ACTIVITY OF AQUALYSIN I ITSELF, IN THAT ORDER. THE
 C-TERMINAL PRO-SEQUENCE IS REQUIRED FOR TRANSLLOCATION OF THE
 PROTEASES ACROSS THE OUTER MEMBRANE.
 CC -1- PM: TWO DISULFIDE BONDS ARE PRESENT.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 SUBTILASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D90108; BAA14135.1; -;
 DR EMBL: X07734; CA30559.1; -;
 DR PIR: S00620; S00620.
 DR PIR: A35742; A35742.
 DR HSP: P06873; 3PRK.
 DR MEROPS: S08_051; -;
 DR InterPro: IPR000209; Peptidase_S8.
 DR Pfam: PF00082; Peptidase_S8; 1.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PROSITE: PS00136; SUBTILASE_ASP; 1.
 DR PROSITE: PS00137; SUBTILASE_HIS; 1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 DR HydroLase: Serine protease; Zymogen; Signal.
 FT SIGNAL 1 14
 FT PROPEP 15 127
 FT CHAIN 128 408 AQUALYSIN I.
 FT PROPEP 409 513
 FT ACT_SITE 166 166 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 197 197 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 349 349 CHARGE RELAY SYSTEM (BY SIMILARITY).
 SQ SEQUENCE 513 AA; 53913 MW; DDFDF6DA50B785 CRC64;
 Query Match 77.8%; Score 28; DB 1; Length 513;
 Best Local Similarity 50.0%; Pred. No. 47;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 wkxxayxg 8
 Db 495 WRIVAYSG 502
 RESULT 18
 HMDH_PICUA
 ID HMDH_PICUA STANDARD; PRT; 934 AA.
 AC 074164;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DE 20-AUG-2001 (Rel. 40, Last annotation update)
 DE 3-HYDROXY-3-METHYLDLUTARYL-COENZYME A REDUCTASE (EC 1.1.1.34) (HMG-COA
 DE REDUCTASE).
 GN HMG.
 OS Pichia jadinii (Yeast) (Candida utilis).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Pichia.
 OX NCBI_TaxId=4903;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98320604; PubMed=9647847;
 RA Shimada H., Kondo K., Fraser P.D., Miura Y., Saito T., Misawa N.;
 RT "Increased carotenoid production by the food yeast *Candida utilis*
 RT through metabolic engineering of the isoprenoid pathway.";
 RL Appl. Environ. Microbiol. 64:2676-2680(1998).
 CC -1- FUNCTION: INVOLVED IN THE CONTROL OF CHOLESTEROL BIOSYNTHESIS. IT
 CC IS THE RATE-LIMITING ENZYME OF THE STEROL BIOSYNTHESIS.
 CC -1- CATALYTIC ACTIVITY: (R)-MEVALONATE + COA + 2 NADP(+) = (S)-3-
 CC HYDROXY-3-METHYLGUTARYL-COA + 2 NADPH.
 CC -1- PATHWAY: CHOLESTEROL BIOSYNTHESIS.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
 CC RETICULUM.
 CC -1- SIMILARITY: BELONGS TO THE HMG-COA REDUCTASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AB012603; BAA31937.1; -;
 DR InterPro: IPR002202; HMG-COA_red.
 DR InterPro: IPR00731; HMGCR_patched_5TM.
 DR Pfam: PF00368; HMG-COA_red; 1.
 DR PRINTS: PR00071; HMGCOAREDUCTASE.
 DR PROSITE: PS00066; HMG_COA_REDUCTASE_1; 1.
 DR PROSITE: PS00318; HMG_COA_REDUCTASE_2; 1.
 DR PROSITE: PS01192; HMG_COA_REDUCTASE_3; 1.
 DR PROSITE: PS50065; HMG_COA_REDUCTASE_4; 1.
 DR PROSITE: PS50156; SSD; 1.
 DR Oxidoreductase; Glycoprotein; Endoplasmic reticulum; Transmembrane;
 KW Cholesterol biosynthesis; NADP.
 FT DOMAIN 1 442
 FT DOMAIN 443 521
 FT DOMAIN 522 934
 FT TRANSMEM 112 132 CATALYTIC.
 FT TRANSMEM 142 162 POTENTIAL.
 FT TRANSMEM 257 277 POTENTIAL.
 FT TRANSMEM 335 355 POTENTIAL.
 FT TRANSMEM 422 442 POTENTIAL.
 FT ACT_SITE 618 618 BY SIMILARITY.
 FT ACT_SITE 828 828 BY SIMILARITY.

FT ACT SITE 924 924 GENERAL BASE (BY SIMILARITY).
FT CARBOHYD 361 361 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 364 364 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 682 682 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 934 AA: 101233 MW: 3C3EA3FC5623601C CRC64:

Query Match 77.8%; Score 28; DB 1; Length 934;
Best Local Similarity 50.0%; Pred. No. 84;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 wxxxxxxg 8
| : | : | :
Db 79 WRSNAYHG 86

RESULT 19

VAT_CAMVW 1D VAT_CAMVW STANDARD; PRT; 99 AA.
AC 001087;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE APHID TRANSMISSION PROTEIN (PROTEIN 2) (FRAGMENT).
GN II.
OS Cauliflower mosaic virus (strain W260) (CamV).
OC Viruses; Retroviral viruses; Caulimovirus.
ON NCBI_TaxID=31558;
RX MEDLINE=92391094; PubMed=1519358;
RA Olu S.G., Schoelz J.E.;
RT "Three regions of cauliflower mosaic virus strain W260 are involved
in systemic infection of solanaceous hosts.";
RL Virology 190:773-782(1992).
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN VIRUS TRANSMISSION.
CC -!- SIMILARITY: BELONGS TO THE CAULIMOVIRUSES ORF II FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M94887; AAA46364.1; -;
FT NON_TER 99 99
SQ SEQUENCE 99 AA: 11049 MW: 4E9942058D0B04B8 CRC64:

Query Match 75.0%; Score 27; DB 1; Length 99;
Best Local Similarity 50.0%; Pred. No. 16;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 wxxxxxxg 8
| : | : | :
Db 61 WKINSYFG 68

RESULT 20

VAT_CAMVW 1D VAT_CAMVW STANDARD; PRT; 159 AA.
AC P03549;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE APHID TRANSMISSION PROTEIN (PROTEIN 2).
GN II.
OS Cauliflower mosaic virus (strain CM-1841) (CamV).
OC Viruses; Retroviral viruses; Caulimovirus.
ON NCBI_TaxID=10644;

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82014878; PubMed=6269062;
RA Gardner R.C., Howarth A.J., Hahn P., Brown-Luedi M., Shepherd R.J.,
Messing J.;
RT "The complete nucleotide sequence of an infectious clone of
RT cauliflower mosaic virus by M13mp7 shotgun sequencing.";
RL Nucleic Acids Res. 9:2871-2888(1981).
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN VIRUS TRANSMISSION.
CC -!- SIMILARITY: BELONGS TO THE CAULIMOVIRUSES ORF II FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; V00140; CAA23453.1; -;
DR PIR; A04157; QOCV2.
SQ SEQUENCE 159 AA: 17900 MW: E6643F75B299B450 CRC64;

Query Match 75.0%; Score 27; DB 1; Length 159;
Best Local Similarity 50.0%; Pred. No. 26;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 wxxxxxxg 8
| : | : | :
Db 61 WKINSYFG 68

Search completed: January 14, 2002, 07:40:37
Job time: 503 sec

..

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:39:39 ; Search time 112.89 Seconds
(without alignments)
10.366 Million cell updates/sec

Title: 09-185908-lb

Perfect score: 36

Sequence: 1 wxxxxxxg 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database :
1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_PROTOZOA:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	83.3	166	11 Q99KK2	Q99KK2 mus musculus
2	30	83.3	432	11 Q99KK2	Q99KK2 mus musculus
3	30	83.3	434	4 Q99KK2	Q99KK2 mus musculus
4	30	83.3	434	4 Q99KK2	Q99KK2 mus musculus
5	30	83.3	613	2 Q99KK2	Q99KK2 mus musculus
6	30	83.3	1039	2 Q99KK2	Q99KK2 mus musculus
7	29	80.6	193	11 Q99KK2	Q99KK2 mus musculus
8	29	80.6	479	12 Q99KK2	Q99KK2 mus musculus
9	29	80.6	497	4 Q99KK2	Q99KK2 mus musculus
10	29	80.6	817	3 Q99KK2	Q99KK2 mus musculus
11	29	80.6	1010	3 Q99KK2	Q99KK2 mus musculus
12	29	80.6	1083	3 Q99KK2	Q99KK2 mus musculus
13	29	80.6	6875	6 Q99KK2	Q99KK2 mus musculus
14	29	80.6	26926	4 Q99KK2	Q99KK2 mus musculus
15	29	80.6	125	5 Q99KK2	Q99KK2 mus musculus
16	28	77.8	303	12 Q99KK2	Q99KK2 mus musculus
17	28	77.8	429	2 Q99KK2	Q99KK2 mus musculus
18	28	77.8	432	5 Q99KK2	Q99KK2 mus musculus
19	28	77.8	432	5 Q99KK2	Q99KK2 mus musculus

20	28	77.8	441	2 Q99KK2	Q99KK2 mus musculus
21	28	77.8	493	12 Q99KK2	Q99KK2 mus musculus
22	28	77.8	493	12 Q99KK2	Q99KK2 mus musculus
23	28	77.8	635	5 Q99KK2	Q99KK2 mus musculus
24	28	77.8	4152	2 Q99KK2	Q99KK2 mus musculus
25	28	77.8	4919	2 Q99KK2	Q99KK2 mus musculus
26	27	75.0	64	12 Q99KK2	Q99KK2 mus musculus
27	27	75.0	99	12 Q99KK2	Q99KK2 mus musculus
28	27	75.0	109	2 Q99KK2	Q99KK2 mus musculus
29	27	75.0	159	12 Q99KK2	Q99KK2 mus musculus
30	27	75.0	159	12 Q99KK2	Q99KK2 mus musculus
31	27	75.0	159	12 Q99KK2	Q99KK2 mus musculus
32	27	75.0	159	12 Q99KK2	Q99KK2 mus musculus
33	27	75.0	246	5 Q99KK2	Q99KK2 mus musculus
34	27	75.0	260	2 Q99KK2	Q99KK2 mus musculus
35	27	75.0	261	2 Q99KK2	Q99KK2 mus musculus
36	27	75.0	262	2 Q99KK2	Q99KK2 mus musculus
37	27	75.0	285	2 Q99KK2	Q99KK2 mus musculus
38	27	75.0	288	2 Q99KK2	Q99KK2 mus musculus
39	27	75.0	303	10 Q99KK2	Q99KK2 mus musculus
40	27	75.0	369	2 Q99KK2	Q99KK2 mus musculus
41	27	75.0	369	2 Q99KK2	Q99KK2 mus musculus
42	27	75.0	369	2 Q99KK2	Q99KK2 mus musculus
43	27	75.0	377	5 Q99KK2	Q99KK2 mus musculus
44	27	75.0	377	5 Q99KK2	Q99KK2 mus musculus
45	27	75.0	377	5 Q99KK2	Q99KK2 mus musculus
46	27	75.0	377	5 Q99KK2	Q99KK2 mus musculus
47	27	75.0	378	2 Q99KK2	Q99KK2 mus musculus
48	27	75.0	379	10 Q99KK2	Q99KK2 mus musculus
49	27	75.0	396	10 Q99KK2	Q99KK2 mus musculus
50	27	75.0	396	10 Q99KK2	Q99KK2 mus musculus

ALIGNMENTS

RESULT 1					
ID Q99KK2	PRELIMINARY:	PRT:	166 AA.		
AC Q99KK2					
DT 01-JUN-2001 (TREMBLrel. 17, Created)					
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)					
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)					
DE SIMILAR TO CYTIDINE MONOPHOSPHO-N-ACETYLTRANSFERASE					
DE (FRAGMENT).					
OS Mus musculus (Mouse).					
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
OX NCBI_TaxID=10090;					
RN [1]					
RP SEQUENCE FROM N.A.					
RA Strausberg R.;					
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.					
DR EMBL: BC004606; AA04606.1; -					
FT NON_TER					
SO SEQUENCE	166 AA;	18201 MW;	B70B2B32F38DA3D8 CRC64;		
Query Match	83.3%;	Score 30;	DB 11;	Length 166;	
Best Local Similarity	62.5%;	Pred. No. 25;			
Matches	5;	Conservative	0;	Mismatches	3;
				Indels	0;
				Gaps	0;
QY 1 wxxxxxxg 8					
Db 97 WKFAVILG 104					
RESULT 2					
ID Q99KK2	PRELIMINARY:	PRT:	432 AA.		
AC Q99KK2					
DT 01-NOV-1998 (TREMBLrel. 08, Created)					
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)					

```

DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE CMP-N-ACETYLNEURAMINIC ACID SYNTHETASE (EC 2.7.7.43) (ACYLNEURAMINATE
DE CYTIDYLYLTRANSFERASE) (CMP-SIALATE PYROPHOSPHORYLASE) (CMP-SIALATE
DE SYNTHASE).
GN CMAS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98356124; Pubmed=9689047;
RA Munster A.K., Eckhardt M., Potvin B., Muhlendorff M., Stanley P.,
RA Gerardy-Schahn R.;
RT "Mammalian cytidine 5'-monophosphate N-acetylneuraminic acid
RT synthetase: a nuclear protein with evolutionarily conserved structural
RT motifs.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:9140-9145(1998).
CC -I- CATALYTIC ACTIVITY: CTP + N-ACYLNEURAMINATE = PYROPHOSPHATE +
CC CMP-N-ACYLNEURAMINATE.
CC EMBL: AJ006215; CAA06915.1; -.
DR MGD; MGI:1337124; Cmas.
DR InterPro; IPR003329; Cytidylyl_trans.
DR Pfam; PF02348; Cytidylyl_trans; 1.
KW Transferase; Nucleotidyltransferase.
SQ SEQUENCE 432 AA; 48058 MW; BF7295535E7F6C65 CRC64;

Query Match      83.3%; Score 30; DB 11; Length 432;
Best Local Similarity 62.5%; Pred. No. 70;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxayxg 8
Db 363 WKEVAVLG 370

RESULT 3
O9NOZ0 PRELIMINARY; PRT; 434 AA.
AC O9NOZ0;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE CMP-N-ACETYLNEURAMINIC ACID SYNTHASE (EC 2.7.7.43).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX Bouguin T., Mundy J.;
RT "Human mRNA for CMP-N-acetylneuraminic acid synthase.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX Iisogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RX Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RX Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RX Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RX Wakamatsu A., Nakamura Y., Nagahari K., Masubo Y., Oshima A.;
RT "NDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN EMBL; AF271386; AAF76203.1; -.
DR EMBL; AK022927; BAB14311.1; -.
DR InterPro; IPR003329; Cytidylyl_trans.
DR Pfam; PF02348; Cytidylyl_trans; 1.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_1.
KW Transferase; Nucleotidyltransferase.
SQ SEQUENCE 434 AA; 48379 MW; B303B647EF81A3A CRC64;

```

```

Query Match      83.3%; Score 30; DB 4; Length 434;
Best Local Similarity 62.5%; Pred. No. 70;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxayxg 8
Db 365 WKEVAVLG 372

RESULT 4
O91IE5 PRELIMINARY; PRT; 538 AA.
AC O91IE5;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE PROBABLE SULFATASE.
GN PA2333.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20437337; Pubmed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Latbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Iori S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004659; AAG05721.1; -.
DR InterPro; IPR000917; Sulfatase.
DR Pfam; PF00884; Sulfatase; 1.
KW Complete proteome.
SQ SEQUENCE 538 AA; 60800 MW; 04CBA9BBE941D340 CRC64;

Query Match      83.3%; Score 30; DB 2; Length 538;
Best Local Similarity 62.5%; Pred. No. 88;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxayxg 8
Db 455 WKYIAYDG 462

RESULT 5
O9A956 PRELIMINARY; PRT; 613 AA.
AC O9A956;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE TONB-DEPENDENT RECEPTOR, PUTATIVE.
GN CC1138.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21173698; Pubmed=11259647;
RA Niernan W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,

```

RA Salberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of *Caulobacter crescentus*."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL: AE005792; AAK23122.1; -
DR TIGR: CC1138; -
KW Receptor: Complete proteome.
SQ SEQUENCE 613 AA; 64411 MW; ED65208A3D82B97D CRC64;

Query Match 83.3%; Score 30; DB 2; Length 613;
Best Local Similarity 50.0%; Pred. No. 1e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxyxg 8
| : | | |
DB 370 WRAAYVSG 377

RESULT 6
ID 09ACP5 PRELIMINARY; PRT; 1039 AA.
AC 09ACP5;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PUTATIVE LARGE SECRETED PROTEIN.
GN SCPI.273C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID:1902;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Brown S.P., Murphy L.D., Harris D.;
RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Bentley S.D., Parkhill J., Barrett B.G., Rajandream M.A.;
RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
RN 13
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA MDDLIN:98241550; PubMed:9573173;
RA Redenbach M., Ikeda K., Yamasaki M., Kinashi H.;
RT "Cloning and physical mapping of the *EcoRI* fragments of the giant
linear plasmid SCPI."
RL J. Bacteriol. 180:2796-2799(1998).
EMBL: AL590464; CAC36799.1; -
SQ SEQUENCE 1039 AA; 110123 MW; ABCA1F4013696B7A CRC64;

Query Match 83.3%; Score 30; DB 2; Length 1039;
Best Local Similarity 62.5%; Pred. No. 1.8e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxyxg 8
| : | | |
DB 603 WKVAYVSG 610

RESULT 7
ID 09ET38 PRELIMINARY; PRT; 193 AA.
AC 09ET38;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CLAUDIN-19 (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID:10090;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-ICR;
RA Kiuchi Y., Morita K., Furuse M., Tsukita S.;
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF249888; AAF96323.1; -
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin; 1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 193
FT NON_TER 193
SQ SEQUENCE 193 AA; 20299 MW; 2F2D82DB5FCF0D7F CRC64;

Query Match 80.6%; Score 29; DB 11; Length 193;
Best Local Similarity 50.0%; Pred. No. 50;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxyxg 8
| : | | |
DB 20 WKQSSVYG 27

RESULT 8
ID 09QON6 PRELIMINARY; PRT; 479 AA.
AC 09QON6;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PUTATIVE APHID TRANSMISSION FACTOR (FRAGMENT).
OS Sugarcane yellow leaf virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
OC Unassigned Luteoviridae.
OX NCBI_TaxID:94290;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-A;
RA Moonan F., Molina J.J., Mirkov T.E.;
RT "Sugarcane yellow leaf virus is a new virus with a genome that has
poliovirus, luteovirus, and Enamovirus properties."
RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
EMBL: AF157029; AAD45687.1; -
DR EMBL: AF157029; AAD45687.1; -
DR InterPro: IPR000893; Luteo_ORF6.
DR InterPro: IPR002929; PLRV_ORF5.
DR InterPro: IPR002965; P-rich_extensn.
DR Pfam: PF01690; PLRV_ORF5; 1.
DR PRINTS: PR00910; LVIRUSORF6.
DR PRINTS: PR01217; PRICHEXTENSN.
FT NON_TER 1
SQ SEQUENCE 479 AA; 51676 MW; 6C1DBAA2BA5EEC9 CRC64;

Query Match 80.6%; Score 29; DB 12; Length 479;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxyxg 8
| : | | |
DB 129 WRAVYVSG 136

RESULT 9
ID 09JH75 PRELIMINARY; PRT; 479 AA.
AC 09JH75;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PUTATIVE APHID TRANSMISSION FACTOR (FRAGMENT).

OS sugarcane yellow leaf virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
 OC Unassigned Luteoviridae.
 OX NCBI_TaxID=94290;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. CP65-357;
 RA Smith G.R.;
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. CP65-357;
 RA MEDLINE=20318675; PubMed=10859394;
 RT "Sugarcane yellow leaf virus: a novel member of the Luteoviridae that probably arose by inter-species recombination."
 RL J. Gen. Virol. 81:1865-1869(2000).
 DR EMBL; AJ249447; CAB75437.1; -
 DR InterPro; IPR000893; Luteo_ORF6.
 DR InterPro; IPR002929; PLRV_ORF5.
 DR Pfam; PF01690; PLRV_ORF5.1.
 DR PRINTS; PRO0910; LVIRUSORF6.
 DR PRINTS; PRO1217; PRICHEXTENSN.
 FT NON_TER 1
 SQ SEQUENCE 479 AA; 51616 MW; AD3B98A1B558652F CRC64;

Query Match 80.6%; Score 29; DB 12; Length 479;
 Best Local Similarity 50.0%; Pred. No. 1.3e+02;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 OY 1 wkxxayxg 8
 | | | |
 Db 129 WRAYAYSG 136

RESULT 10
 O9BRR6 PRELIMINARY; PRT; 497 AA.
 AC O9BRR6;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE SIMILAR TO RIKEN CDNA 2610017G09 GENE.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=KIDNEY ADENOCARCINOMA;
 RA Strausberg R.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC006112; AAH06112.1; -
 SQ SEQUENCE 497 AA; 54088 MW; B758E977CDA88F8F CRC64;

Query Match 80.6%; Score 29; DB 4; Length 497;
 Best Local Similarity 50.0%; Pred. No. 1.4e+02;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 OY 1 wkxxayxg 8
 | | | |
 Db 4 WRGSAYAG 11

RESULT 11
 O9P7A3 PRELIMINARY; PRT; 817 AA.
 AC O9P7A3;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE UBIQUITIN FUSION DEGRADATION PROTEIN-2 (FRAGMENT).
 GN UFD2.
 OS Schizosaccharomyces pombe. (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972H-;
 RA Saunders D., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL163640; CAB87206.1; -
 FT NON_TER 817 817
 SQ SEQUENCE 817 AA; 93232 MW; AE3B4CC2E118279B CRC64;

Query Match 80.6%; Score 29; DB 3; Length 817;
 Best Local Similarity 62.5%; Pred. No. 2.3e+02;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 OY 1 wkxxayxg 8
 | | | |
 Db 706 WKOPAYFG 713

RESULT 12
 O9HE05 PRELIMINARY; PRT; 1010 AA.
 AC O9HE05;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE UBIQUITIN FUSION DEGRADATION PROTEIN-2.
 GN UFD2.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972H-;
 RA Seeger K., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL51487; CAC19740.1; -
 DR InterPro; IPR003613; Ubpx.
 DR SMART; SM00504; Ubpx; 1.
 SQ SEQUENCE 1010 AA; 115276 MW; 60A3E5B5E790FE6E CRC64;

Query Match 80.6%; Score 29; DB 3; Length 1010;
 Best Local Similarity 62.5%; Pred. No. 2.9e+02;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 OY 1 wkxxayxg 8
 | | | |
 Db 706 WKOPAYFG 713

RESULT 13
 O9RPH6 PRELIMINARY; PRT; 1083 AA.
 AC O9RPH6;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE RECB.
 GN RECB.
 OS Mycobacterium smegmatis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

```
OX NCBI_TaxID:1772;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC2155 (NB2);
RX MEDLINE=99412429; PubMed:10481025;
RA Griffin T.J. IV, Parsons L., Leschziner A.E., Devost J.,
  "In vitro transposition of Tn552: a tool for DNA sequencing and
  mutagenesis.";
RL Nucleic Acids Res. 27:3859-3865(1999).
DR HSSP; P56255; 2PJR.
DR Interpro: IPR001993; Mitoch_carrier.
DR Interpro: IPR00212; Uvrd_hellcase.
DR Pfam: PF00580; Uvrd-hellcase; 1.
DR PROSITE: PS00215; MITOCH_CARRIER; UNKNOWN; 1.
SQ SEQUENCE 1083 AA; 117442 MW; A531F5FBA5EB748 CRC64;

Query Match      80.6%; Score 29; DB 2; Length 1083;
Best Local Similarity 50.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxxg 8
   |||||
Db 766 WKRTSYSG 773

RESULT 14
ID Q28733 PRELIMINARY; PRT; 6875 AA.
AC Q28733; Q28736;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE TITIN (FRAGMENT).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID:9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CE12;
RA Labelt S.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-6805 FROM N.A.
RC STRAIN=CE12;
RX MEDLINE=92258380; PubMed:1582406;
RA Labelt S., Gautel M., Lakey A., Trinick J.;
RT "Towards a molecular understanding of titin.";
RL EMBO J. 11:1711-1716(1992).
RN [3]
RP SEQUENCE OF 4305-5320 FROM N.A.
RC TISSUE=PSOAS MUSCLE;
RX MEDLINE=90238553; PubMed:2129545;
RA Labelt S., Barlow D.P., Gautel M., Gibson T., Holt J., Hsieh C.L.,
  Francke U., Leonard K., Martale J., Whiting A., Trinick J.;
RT "A regular pattern of two types of 100-residue motif in the sequence
  of titin.";
RL Nature 345:273-276(1990).
CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
  DOMAIN.
CC EMBL: X64696; CAA45937.1; -.
CC HSSP; P56276; ITLK.
CC Interpro: IPR000282; CytoK_receptor_2.
CC Interpro: IPR003962; FnIII_repeat.
CC Interpro: IPR003961; FN_III.
CC Interpro: IPR003598; Ig_c2.
CC Interpro: IPR003600; Ig_like.
CC Interpro: IPR003006; Ig_MHC.
DR Pfam: PF00041; fn3; 50.
```

```
DR Pfam: PF00047; Ig; 15.
DR PRINTS; PR00014; ENTPEP11.
DR SMART; SM00060; FN3; 48.
DR SMART; SM00408; IGC2; 3.
DR SMART; SM00410; IG_Like; 15.
KW Muscle protein; Myosin; Repeat.
FT NON_TER 1
FT NON_TER 6875
SQ SEQUENCE 6875 AA; 759127 MW; 50CA5B84F3668C55 CRC64;

Query Match      80.6%; Score 29; DB 6; Length 6875;
Best Local Similarity 62.5%; Pred. No. 2.2e+03;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxxg 8
   |||||
Db 6555 WKRPAYDG 6562

RESULT 15
ID Q10466 PRELIMINARY; PRT; 26926 AA.
AC Q10466;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE TITIN, HEART ISOFORM N2-B (EC 2.7.1.-) (CONNECTIN).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEART;
RX MEDLINE=96026330; PubMed-7569978;
RA Labelt S., Kolmer B.;
RT "Titlins: giant proteins in charge of muscle ultrastructure and
  elasticity.";
RL Science 270:293-296(1995).
RN [2]
RP SEQUENCE OF 22277-25376 FROM N.A.
RX MEDLINE=92258380; PubMed:1582406;
RA Labelt S., Gautel M., Lakey A., Trinick J.;
RT "Towards a molecular understanding of titin.";
RL EMBO J. 11:1711-1716(1992).
RN [3]
RP SEQUENCE OF 1976-2014 FROM N.A.
RA Labelt S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=95331314; PubMed-7607248;
RA Gautel M., Castiglione-Morelli M.A., Pfehl M., Motta A., Pastore A.;
RT "A calmodulin-binding sequence in the C-terminus of human cardiac
  titin kinase.";
RL Eur. J. Biochem. 230:752-759(1995).
CC -1- FUNCTION: THIS GIANT MUSCLE PROTEIN MAY BE INVOLVED IN MUSCLE
  ASSEMBLY AND IN MAINTAINING THE STRUCTURAL INTEGRITY OF
  SARCOMERES. MAY HAVE PROTEIN KINASE ACTIVITY.
CC -1- ALTERNATIVE PRODUCTS: A NUMBER OF FORMS OF THIS PROTEIN ARE
  PRODUCED BY ALTERNATIVE SPLICING WHICH DIFFER IN TISSUE
  DISTRIBUTION. DIFFERENT SIZE TRANSCRIPTS MAY ALSO EXIST WITHIN ANY
  ONE TISSUE. THE SEQUENCE SHOWN HERE IS THAT OF THE HEART ISOFORM
  N2-B.
CC -1- TISSUE SPECIFICITY: MUSCLE-SPECIFIC.
CC -1- SIMILARITY: TO THE CATALYTIC DOMAINS OF OTHER SERINE/THREONINE
  KINASES.
CC -1- SIMILARITY: BELONGS TO IMMUNOGLOBULIN SUPERFAMILY. CONTAINS 112
  IMMUNOGLOBULIN C2-LIKE DOMAINS AND 132 FIBRONECTIN TYPE III-LIKE
  DOMAINS.
CC EMBL: X64698; CAA45939.1; -.
DR EMBL: X83270; CAA58243.1; -.
```

DR EMBL: X64697; CAA45938.1; -
 DR EMBL: X90568; CAA62188.1; -
 DR EMBL: X64699; CAA45940.1; -
 DR HSSP: P56276; ITTK.
 DR InterPro: IPR000282; CytoK_receptor_2.
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR000577; FGfy_kin.
 DR InterPro: IPR003962; FhII_repea.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR003015; HLH_Myc.
 DR InterPro: IPR003598; Ig_C2.
 DR InterPro: IPR003600; Ig_C2.
 DR InterPro: IPR003006; Ig_EHC.
 DR InterPro: IPR000129; Peroxidase.
 DR InterPro: IPR002016; Peroxidase.
 DR InterPro: IPR002290; Ser_thr_kin_actsite.
 DR InterPro: IPR001245; Tyr_kin.
 DR Pfam: PF00041; fn3; 132.
 DR Pfam: PF00047; Ig; 95.
 DR Pfam: PF00069; pkinase; 1.
 DR PRINTS: PR00014; ENTPERTI.
 DR PRINTS: PR00726; LEXASERPTASE.
 DR SMART: SM00060; FN3; 127.
 DR SMART: SM00408; IgC2; 23.
 DR SMART: SM00410; Ig_Like; 79.
 DR SMART: SM00220; S_TKC; 1.
 DR PROSITE: PS00933; FGfy_KINASES_1; UNKNOWN_1.
 DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
 DR PROSITE: PS00290; IG_MHC; UNKNOWN_1.
 DR PROSITE: PS00435; PEROXIDASE_1; UNKNOWN_1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 KW Muscle protein; Cytoskeleton; Structural protein; Calmodulin-binding;
 KW Serine/threonine-protein kinase; Alternative splicing; Repeat;
 KW Immunoglobulin domain; Phosphorylation.
 FT DOMAIN 1370 1389
 FT DOMAIN 4 5 AA TANDEM REPEATS OF R-M-S-P-A.
 FT DOMAIN 429 4614
 FT DOMAIN 24731 25070
 FT DOMAIN 25030 25056
 FT MOD_RES 1372 1372
 FT MOD_RES 1377 1377
 FT MOD_RES 1382 1382
 FT MOD_RES 1387 1387
 FT MOD_RES 26171 26171
 FT MOD_RES 26178 26178
 FT MOD_RES 26184 26184
 FT MOD_RES 26190 26190
 FT MOD_RES 22277 22277
 FT CONFLICT 22449 22449
 FT CONFLICT 22454 22454
 FT CONFLICT 23324 23324
 FT CONFLICT 23324 23324
 SQ SEQUENCE 26926 AA; 2993428 MW; D5EECD3254DF5523 CRC64;

Query Match 80.6%; Score 29; DB 4; Length 26926;
 Best Local Similarity 62.5%; Pred. No. 9.1e+03;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxyg 8
 DB 18126 WKPPAYDG 18133

RESULT 16

ID O9NKN6 PRELIMINARY; PRT; 125 AA.
 AC O9NKN6;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
 DE IRON TRANSPORT MULTICOPPER OXIDASE RELATED PROTEIN 1 (FRAGMENT).
 GN ITMCRP1.
 OS Leishmania major.

OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5664;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FRIEDLIN;
 RA Myler P.J., Sisk E., Hixson G., Kiser P., Rickel E., Hassebrock M.,
 RA Cawthra J., Marsolini F., Sunkin S., Stuart K.D.;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC005935; AAF39743.1; -
 DR InterPro: IPR002355; Multicu_oxidase2.
 DR PROSITE: PS00080; MULTICOPPER_OXIDASE2; UNKNOWN_1.
 FT NON_TER 1
 SQ SEQUENCE 125 AA; 13896 MW; C017C9416C7DAE31 CRC64;

Query Match 77.8%; Score 28; DB 5; Length 125;
 Best Local Similarity 50.0%; Pred. No. 53;
 Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wxxxxxyg 8
 DB 9 WKATYVG 16

RESULT 17
 ID O84711 PRELIMINARY; PRT; 303 AA.

AC O84711;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
 DE ORF 1, COMPLETE CDS; ORF 2, COMPLETE CDS; ORF 3, 3' END CDS; COAT
 DE PROTEIN (FRAGMENT).
 OS Pea enation mosaic virus (PEMV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
 OC Enamovirus.
 OX NCBI_TaxID=12290;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MSG.
 RX MEDLINE=91341468; Pubmed=1875194;
 RA Demler S.A., de Zoeten G.A.;
 RT "The nucleotide sequence and luteovirus-like nature of RNA 1 of an
 RT aphid non-transmissible strain of pea enation mosaic virus";
 RL J. Gen. Virol. 72:1819-1834(1991).
 DR EMBL: L04573; AAF7298.1; -
 DR InterPro: IPR000893; Luteo_ORF6.
 DR InterPro: IPR002929; PLRV_ORF5.
 DR Pfam: PF01690; PLRV_ORF5; 1.
 DR PRINTS: PR00910; LIVIRSORF6.
 FT NON_TER 1
 SQ SEQUENCE 303 AA; 33449 MW; F20BA2D9D72A5AE2 CRC64;

Query Match 77.8%; Score 28; DB 12; Length 303;
 Best Local Similarity 50.0%; Pred. No. 1.4e+02;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxxyg 8
 DB 135 WQARAYTG 142

RESULT 18

ID O9RDP7 PRELIMINARY; PRT; 429 AA.
 AC O9RDP7;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
 DE BETA-KETOACYL SYNTHASE.
 GN FABB.
 OS Streptomyces coelicolor.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1902;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN (12)
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN (13)
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RC MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleiser H.M., Denapalte D., Elchner A., Cullum J.,
RA Kinusch H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL: AL134423; CAB62722.1; -;
DR HSSP: P39435; 1KAS;
DR InterPro: IPR000794; Ketoacyl-synt.
DR Pfam: PF00109; Ketoacyl-synt. 1.
SQ SEQUENCE 429 AA: 43841 MW: 983DC0F9800582BE CRC64;

Query Match 77.8% Score 28; DB 2; Length 429;
Best Local Similarity 50.0%; Pred. NO. 2e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wxxxxyxg 8
|| | | |
DB 87 WKDAGYNG 94

RESULT 19
ID 09NFP2 PRELIMINARY; PRT: 432 AA.
AC 09NFP2;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE NIMA-RELATED PROTEIN KINASE (FRAGMENT).
GN NEK-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN (1)
RP SEQUENCE FROM N.A.
RA Doria D., Le Roch K., Scallafandro P., Alano P., Parzy D., Meijer L.,
RA Doerig C.;
RT "Pfnck-1, a novel NIMA-related protein kinase from the human malaria
parasite Plasmodium falciparum.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL: AJ276023; CAB76949.1; -;
DR InterPro: IPR000719; Euk_Pkinase.
DR InterPro: IPR002290; Ser_Thr_Kin_actsite.
DR Pfam: PF00069; S_TKc; 1.
DR SMART: SM00220; S_TKc; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
FT NON-TER 432 432
SQ SEQUENCE 432 AA: 49698 MW: C5FFD6A7C929E5D CRC64;

Query Match 77.8% Score 28; DB 5; Length 432;
Best Local Similarity 50.0%; Pred. NO. 2e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxyxg 8
|| | | |
DB 43 WKAIYRG 50

RESULT 20
ID 060035 PRELIMINARY; PRT: 441 AA.
AC 060035;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE 4-ALPHA-GLUCANOTRANSFERASE (EC 2.4.1.25) (DISPROPORTIONATING ENZYME)
DE (D-ENZYME) (DEXTRIN GLYCOSYLTRANSFERASE) (OLIGO-1,4-1,4-
DE GLUCANTRANSFERASE).
GN MGT.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-MSB8 (DSM3109);
RC MEDLINE=92331687; PubMed=1628664;
RA Liebl W., Fell R., Gabelsberger J., Kellermann J., Schleifer K.;
RT "Purification and characterization of a novel thermostable 4-alpha-
glucanotransferase of Thermotoga maritima cloned in Escherichia
coli.";
RL Eur. J. Biochem. 207:81-88(1992).
RN (12)
RP SEQUENCE FROM N.A.
RC STRAIN-MSB8 (DSM3109);
RC Heinrich P., Huber W.;
RL Syst. Appl. Microbiol. 17:297-305(1994).
RN (13)
RP SEQUENCE FROM N.A.
RC STRAIN-MSB8 (DSM3109);
RA Liebl W.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: TRANSFERS A SEGMENT OF A 1,4-ALPHA-D-GLUCAN TO
A NEW 4-POSITION IN AN ACCEPTOR, WHICH MAY BE GLUCOSE OR 1,4-
ALPHA-D-GLUCAN.
CC EMBL: Z50813; CA90693.1; -;
DR InterPro: IPR00461; Alpha_amylase.
DR Pfam: PF00128; alpha-amylase; 1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 441 AA: 51857 MW: 3D221B5234BAFBAC CRC64;

Query Match 77.8% Score 28; DB 2; Length 441;
Best Local Similarity 62.5%; Pred. NO. 2e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxyxg 8
|| | | |
DB 348 WKMPAYNG 355

Search completed: January 14, 2002, 07:39:41
Job time: 952 sec

XX WPI: 2001-376931/40.
DR N-PSDB; AAH65652.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
PS Claim 17; SEQ ID NO: 4187; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 275 AA;

Query Match 86.1%; Score 31; DB 22; Length 275;
Best Local Similarity 62.5%; Pred. No. 59;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 wkxxxyxg 8
|| || |
Db 130 wkdlayag 137

RESULT 2

AAB28673
ID AAB28673 standard; Protein; 434 AA.

XX AAB28673;

DT 13-FEB-2001 (first entry)

XX Human carbohydrate-modifying enzyme Incyte ID No: 000422CD1.

XX Human; carbohydrate-modifying enzyme; CME; antidiabetic;
KW immunosuppressive; anti-HIV; antiinflammatory; antianaemic;
KW antiasthmatic; antiarteriosclerotic; antihypoid; hepatotropic;
KW nephrotropic; antigout; thyromimetic; neuroprotective; osteopathic;
KW antiarthritic; antipsoriatic; uropathic; ophthalmological;
KW dermatological; antilucer; cytoslatic; vitucide; antibacterial;
KW fungicide; protozoacide; tranquilliser; vunerary; diabetes;
KW autoimmune disorder; inflammatory disorder; infection.

XX Homo sapiens.

XX WO200063351-A2.

XX 26-OCT-2000.

XX 20-APR-2000; 2000WO-US10882.

XX 21-APR-1999; 99US-0130383.

XX (INCY-) INCYTE GENOMICS INC.

XX Lal P, Yue H, Tang YT, Hillman JL, Baughn MR, Yang J;

XX WPI: 2000-672729/65.

XX N-PSDB; AAC65392.

PT Novel carbohydrate modifying enzyme polypeptides and polynucleotides
PT for diagnosis, treatment, and prevention of carbohydrate metabolism
PT disorders, autoimmune/inflammatory disorders, and cancer
PT
XX
PS Claim 1; Page 66-67; 75pp; English.
XX
CC The present sequence is a human carbohydrate-modifying enzyme
CC (CME). CME polynucleotides are useful for treating and
CC diagnosing diseases associated with CME such as diabetes,
CC autoimmune/inflammatory disorders such as AIDS, Addison's disease,
CC adult respiratory distress syndrome, allergies, anaemia, asthma,
CC atherosclerosis, autoimmune thyroiditis, bronchitis, cholecystitis,
CC contact dermatitis, Crohn's disease, emphysema, erythroblastosis fetalis,
CC glomerulonephritis, Good pasture's syndrome, gout, Grave's disease,
CC Hashimoto's thyroiditis, multiple sclerosis, myasthenia gravis,
CC osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis,
CC Reiter's syndrome, arthritis, scleroderma, Sjogren's syndrome, systemic
CC lupus erythematosus, ulcerative colitis, uveitis, Werner syndrome,
CC complications of cancer, haemodialysis, and extracorporeal circulation,
CC viral, bacterial, fungal parasitic, protozoal, and helminthic infections,
CC trauma, or cancer. CME, or its catalytic or immunogenic fragment, is
CC useful for drug screening.
XX
SQ Sequence 434 AA;

Query Match 83.3%; Score 30; DB 21; Length 434;
Best Local Similarity 62.5%; Pred. No. 15e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 wkxxxyxg 8
|| || |
Db 365 wkewaylg 372

RESULT 3

AAB43165
ID AAB43165 standard; Protein; 434 AA.

XX AAB43165;

DT 08-FEB-2001 (first entry)

XX Human ORFX ORF2929 polypeptide sequence SEQ ID NO:5858.

XX Human; open reading frame; ORFX; detection; cytoslatic; hepatotropic;
KW vunerary; antipsoriatic; antiparkinsonian; noctropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antineumatic; antihypoid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.

XX Homo sapiens.

XX WO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

XX 02-APR-1999; 99US-0127636.

XX 05-APR-1999; 99US-0127728.

XX 30-MAR-2000; 2000US-0540763.

PA (CURA-) CURAGEN CORP.
XX
XX Shinketsu RA, Leach M;
XX
DR WPI: 2000-602362/57.
DR N-PSDB: AAC77374.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 11: Page 5028-5029; 5507pp: English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antiproliferic; antiparkinsonian; nootropic; neuroprotective;
CC osteoplastic; anticonvulsant; antitachytic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antitubercular; antibacterial; antiviral; antifungal; antineumatic;
CC antihypertoid; and antianemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 434 AA:
OY 1 wkxxayxg 8
DB 365 wkexaylg 372

Query Match 83.3%; Score 30; DB 21; Length 434;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

RESULT 4
AAV96100 standard; Protein: 434 AA.
XX
XX AAV96100;
XX
XX 19-DEC-2000 (first entry)
XX
XX Human CMP-sialic acid synthetase.
XX
XX CMP-sialic acid synthetase: human; sialylation; glycoprotein;
KW plasmidogen; transferrin; thyrotropin; Na⁺,K⁺-ATPase.
XX
OS Homo sapiens.
XX
XX
XX Key Location/Qualifiers
FH Misc-difference 133 /note= "encoded by YAT"
FT Misc-difference 136 /note= "encoded by GKT"
FT Misc-difference 147 /note= "encoded by YCA"
FT Misc-difference 169 /note= "encoded by GKT"
XX
XX WO200052135-A2.
PN

XX
PD 08-SEP-2000.
XX
XX 01-MAR-2000; 2000WO-US05313.
XX
XX WPI: 2000-572178/53.
DR N-PSDB: AAA50568.
XX
XX Recombinant production of sialylated glycoproteins using cells in which
PT the expression of enzymes, e.g. sialic acid synthetase, involved in the
PT sialylation reaction has been altered -
XX
PS Claim 20; Page 102-103; 144pp: English.
XX
XX The present sequence is that of human CMP-sialic acid synthetase.
CC The invention provides methods and recombinantly engineered cells
CC for producing glycoproteins having sialylated oligosaccharides.
CC The methods involve altering the expression of enzymes involved in
CC carbohydrate processing. A claimed cell producing sialylated
CC glycoprotein at above endogenous levels expresses at least 1
CC (preferably human) enzyme selected from G1CNAc-2 epimerase, an
CC enzyme catalyzing the conversion of UDP-GlcNAc to MannNAc, sialic
CC acid synthetase, aldolase, CMP-sialic acid synthetase and
CC N-acetylglucosaminidase activity may be suppressed. A claimed
CC method for manipulating glycoprotein in an insect cell comprises
CC enhancing the expression of 1 of the above enzymes, and a claimed
CC method for producing sialylated glycoproteins involves expressing
CC a heterologous protein (especially plasmidogen, transferrin,
CC Na⁺,K⁺-ATPase or thyrotropin) in the insect cell. Yeast, insect,
CC fungal, plant and bacterial host cells can be engineered to produce
CC new forms of sialylated glycoproteins, higher concentrations of
CC sialylated glycoproteins and/or elevated concentrations of donor
CC substrates (e.g. nucleotide sugars) required for sialylation.
XX
SQ Sequence 434 AA:
OY 1 wkxxayxg 8
DB 365 wkexaylg 372

Query Match 83.3%; Score 30; DB 21; Length 434;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

RESULT 5
AAV90351 standard; Protein: 434 AA.
XX
XX AAV90351;
XX
XX 04-DEC-2000 (first entry)
XX
XX Human glycosylation enzyme clone HMLM34 protein sequence.
XX
XX Human; glycosylation enzyme; glycolysis; myoglobinuria; tumour marker;
KW immunotherapy; cosmetic surgery; metabolism; immune system disorder;
KW haematopoietic cell deficiency; blood coagulation disorder; asthma;
KW afibrinogenemia; blood platelet disorder; thrombocytopenia; neoplasia;
KW autoimmune disorder; Addison's disease; multiple sclerosis; purpura;
KW allergic encephalomyelitis; allergic reaction; organ rejection;
KW graft-versus-host disease; inflammation; hyperproliferative disorder;
KW

KW sarcoidosis; infection; gene therapy; CMP sialic acid synthetase.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 133 /note= "encoded by YAT"
FT Misc-difference 136 /note= "encoded by GKT"
FT Misc-difference 147 /note= "encoded by YCA"
FT Misc-difference 169 /note= "encoded by GKT"
FT Misc-difference /note= "encoded by GKT"
XX
PN WO20052136-A2.
XX
XX 08-SEP-2000.
XX
XX 01-MAR-2000; 2000WO-US05325.
XX
XX 02-MAR-1999; 99US-0122409.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Coleman TA:
XX
XX WPI: 2000-572179/53.
XX
XX N-PSDB; AAA37762.
XX
XX
XX New human glycosylation enzymes cytidine 5'-monophosphate sialic acid synthetase, sialic acid synthetase and aldolase and nucleic acids encoding the proteins for treating e.g., immune system disorders, microbial diseases -
XX
XX
XX Claim 12; Page 106-108; 115pp; English.
XX
XX This sequence represents a human glycosylation enzyme clone of the invention, designated HMLM34. This protein clone is a CMP sialic acid synthetase. The sequences are useful as reagents for the differential identification of the tissues or cell types present in a biological sample, as immunological probes, for treating a disease or condition resulting from under expression of such polypeptide, for the detection and/or treatment of disorders involving aberrant glycolysis, e.g. cramps, myoglobinuria, and as tumour marker and/or immunotherapy targets. They may also be used to differentiate, proliferate and attract cells leading to the regeneration of tissues, to modulate mammalian characteristics (e.g. in cosmetic surgery) or mammalian metabolism affecting catabolism, anabolism processing, utilisation and energy storage, to change a mammal's mental state by influencing and as a food additive or preservative. The proteins can be used to assay protein levels in a sample, as a marker or detector of an immune system disorder, to inhibit cytokine activity, and as a vaccine. They may further be used to treat immune system or of haematopoietic cell deficiencies or disorders, blood coagulation disorders (e.g. afibrinogenaemia), blood platelet disorders (e.g. thrombocytopenia), wounds resulting from trauma or surgery, autoimmune disorders (e.g. Addison's disease, multiple sclerosis, allergic encephalomyelitis), allergic reactions (e.g. asthma), organ rejection, graft-versus-host disease, inflammation, hyperproliferative disorders (e.g. neoplasia, purpura, sarcoidosis), diseases caused by viruses (e.g. hepatitis, meningitis, AIDS), bacteria and fungi (include e.g. tuberculosis, conjunctivitis, sepsis, typhoid, chlamydia, cellulitis), and diseases caused by parasites (e.g. amoebiasis, coccidiosis, leishmaniasis, scabies, malaria, toxoplasmosis).
XX
SQ Sequence 434 AA:

Query Match 83.3%; Score 30; DB 21; Length 434;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 wkxxxyxg 8
|| || |

DB 365 wkexaylg 372
RESULT 6
ID AAM39319
XX AAM39319 standard; Protein; 434 AA.
XX
XX AAM39319;
XX
XX 22-OCT-2001 (first entry)
XX
XX Human polypeptide SEQ ID NO 2464.
XX
XX Human; nootropic; immunosuppressant; cytotactic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukaemia.
XX
XX Homo sapiens.
XX
XX WO200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AU, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
XX N-PSDB; AAI58475.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries -
XX
XX Example 4; SEQ ID NO 2464; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and the encoded polypeptides (AAM38642-AAM4213) with nootropic, immunosuppressant and cytotactic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed specification.
XX
SQ Sequence 434 AA:

Query Match 83.3%; Score 30; DB 22; Length 434;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;

Matches	5:	Conservative	0:	Mismatches	3:	Indels	0:	Gaps	0:
OY	1	wkxayxg 8							
Db	365	wkevaylg 372							

RESULT 7
AAB84682
ID AAB84682 standard; Protein: 434 AA.
XX
AC AAB84682;
XX
DT 17-SEP-2001 (first entry)
XX
DE Nucleotide sequence of a cytidine monophosphate-sialic acid synthetase.
XX
KM Carbohydrate processing; sialylation; sialylated glycoprotein; CMP-SA;
KW cytidine monophosphate-sialic acid; N-acetylglucosamine-2 epimerase;
KW sialic acid synthetase; aldolase; CMP-SA synthetase; CMP-SA transporter;
vacine.
XX
OS Homo sapiens.
XX
PH Key Location/Qualifiers
FT Misc-difference 133
FT /note= "unspecified residue encoded by YAT"
FT Misc-difference 136
FT /note= "unspecified residue encoded by GKT"
FT Misc-difference 147
FT /note= "unspecified residue encoded by YCA"
FT Misc-difference 169
FT /note= "unspecified residue encoded by GKT"
XX
PN MO200142492-A1.
XX
PD 14-JUN-2001.
XX
PF 07-DEC-2000; 2000MO-US33136.
XX
PR 09-DEC-1999; 99US-0169839.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (UYJO) UNIV JOHNS HOPKINS.
PA (UTEM) UNIV TEMPLE.
PA (UYWY-) UNIV WYOMING.
PI Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA, Palter K, Jarvis D;
DR WPI: 2001-441575/47.
DR N-PSDB; AAH28457.
XX
PT Cells producing cytidine monophosphate-sialic acid and sialylated
PT glycoprotein above endogenous levels for production of vaccines and
PT therapeutics -
XX
PS Claim 20; Fig 30; 182pp; English.
XX
CC The specification describes a method for manipulating carbohydrate
CC processing pathways in cells of interest. The methods are used to
CC manipulate multiple pathways involved with the sialylation reaction by
CC using recombinant DNA technology and substrate feeding approaches to
CC enable the production of sialylated glycoproteins in the cells. The
CC sialylation process involves the post-translational addition of the
CC donor substrate cytidine monophosphate-sialic acid (CMP-SA) onto a
CC specific acceptor carbohydrate. The cells express at least one enzyme,
CC selected from N-acetylglucosamine-2 epimerase, sialic acid synthetase,
CC aldolase, CMP-SA synthetase and CMP-SA transporter, above endogenous
CC levels. The cells are useful for producing complex sialylated
CC glycoproteins in cells of interest, especially insect cells.
CC Glycoproteins containing sialylated oligosaccharides are useful as
CC vaccines, therapeutics and diagnostic tools. Cells producing complex
CC sialylated glycoproteins are useful for enhancing the value of

CC heterologous expression systems and increasing the application of
CC heterologous cell expression products as vaccines, therapeutics and
CC diagnostic tools as well as increasing the variety of heterologous
CC proteins that can be produced and lowering biotechnology production
CC costs. The present sequence represents a human CMP-SA synthetase, which
CC is used in the method of the invention.
XX

SQ Sequence 434 AA;

Query Match 83.3%; Score 30; DB 22; Length 434;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5: Conservative 0: Mismatches 3: Indels 0: Gaps 0:

OY 1 wkxayxg 8
Db 365 wkevaylg 372

RESULT 8
AAB95213
ID AAB95213 standard; Protein: 434 AA.
XX
AC AAB95213;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:17325.
XX
KM Human; primer: detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Negai K, Otsuki T;
XX
DR WPI: 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 17325; 2537pp + CD ROM; English.
XX

CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 434 AA;

SO

Query Match 83.3%; Score 30; DB 22; Length 434;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxayxg 8
|||
Db 365 wkevaylg 372

RESULT 9
AAB43931
ID AAB43931 standard; Protein; 448 AA.

AC AAB43931;
XX
DT 08-FEB-2001 (first entry)

DE Human cancer associated protein sequence SEQ ID NO:1376.

XX
XX Human; cancer associated gene; cancer antigen; detection; cancer;
KM diagnosis; cytostatic; proliferative; vulnerability; immunomodulator;
KM antidiabetic; antiashtmatic; antirheumatic; antiarthritic; anti viral;
KM antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;
KM dermatologic; neuroprotective; thrombolytic; coagulant; nootropic;
KM vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KM immune disorder; haematopoietic cell disorder; autoimmune disorder;
KM allergic reaction; graft versus host disease; organ rejection;
KM haemostatic; thrombolytic; cardiovascular disorder; infection;
KM neurological disease; drug screening.

XX
OS Homo sapiens.
XX
PN WO20005350-A1.
XX
PD 21-SEP-2000.

XX
PF 08-MAR-2000; 2000WO-US05882.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.

XX
PI Rosen CA, Ruben SM;
XX
XX WPI; 2000-587533/55.
DR N-PSDB; AAC78140.
XX

PT Novel isolated nucleic acids comprising sequences encoding peptides
PT useful for treating or diagnosing e.g. cancer -
XX

PS Claim 11; Page 2041-2042; 2352pp; English.

XX
XX AAC77607 to AAC78448 encode the human cancer associated proteins given
CC in AAB43398 to AAB44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; vulnerability; immunomodulator;
CC antidiabetic; antiashtmatic; antirheumatic; antiarthritic;
CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC dermatologic; neuroprotective; cardiac; thrombolytic; coagulant;
CC nootropic; vasotropic; antipsoriatic and antiangiogenic. The
CC polynucleotides and polypeptides can be used for preventing, treating or

CC ameliorating medical conditions and diagnosing pathological conditions.
CC polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells, to treat disorders of haematopoietic cells, autoimmune
CC disorders, allergic reactions, graft versus host disease and organ
CC rejection, modulate haemostatic or thrombolytic activity, modulate
CC inflammation, cancers, cardiovascular disorders, neurological disease and
CC bacterial or viral infections. The peptides, nucleotides, antibodies,
CC agonists and antagonists may be also be used in drug screens. AAC78449 to
CC AAC78457 and AAB44240 represent sequences used in the exemplification of
CC the present invention.

XX Sequence 448 AA;

SO

Query Match 83.3%; Score 30; DB 21; Length 448;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxayxg 8
|||
Db 379 wkevaylg 386

RESULT 10
AAM41105
ID AAM41105 standard; Protein; 461 AA.

AC AAM41105;
XX
DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 6036.

XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KM peripheral nervous system; neuropathy; central nervous system; CNS;
KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KM chemokine; thrombolytic; drug screening; arthritis; inflammation;
KM leukaemia.

XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.

XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
XX
PR 25-APR-2000; 2000US-0552317.
XX
PR 09-JUL-2000; 2000US-0598042.
XX
PR 19-JUL-2000; 2000US-0620312.
XX
PR 03-AUG-2000; 2000US-0653450.
XX
PR 14-SEP-2000; 2000US-0662191.
XX
PR 19-OCT-2000; 2000US-0693036.
XX
PR 29-NOV-2000; 2000US-0727344.

XX
PA (HYSE-) HYSEQ INC.
XX
XX Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Mehman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou F, Goodrich R, Drmanac RT;
XX
XX WPI; 2001-442253/47.
DR N-PSDB; AAI60261.

PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 2; SEQ ID NO 6036; 10078pp; English.

TfE invention relates to human nucleic acids (AA157798-AA161369) and the encoded polypeptides (AA138642-AA142213) with nocotropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, Activin/Inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, Leukaemias and C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed specification.

SO Sequence 461 AA:

Query Match 83.3%; Score 30; DB 22; Length 461;
Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxayxg 8
|||
Db 379 wkxay1g 386

RESULT 11
AA12227 ID AA12227 standard; Protein: 55 AA.

AC AA12227;
XX 18-JUN-1999 (first entry)
DT

DE Human 5' EST secreted protein SEQ ID NO: 540.
XX Human: secreted protein; EST: expressed sequence tag; diagnosis;
XX forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition.
XX Homo sapiens.
OS
XX MO9906554-A2.
PN
XX 11-FEB-1999.
PD
XX 31-JUL-1998; 98MO-1B01238.
PE
XX 01-AUG-1997; 97US-0905134.
PR
XX (GEST) GENSET.
PA
XX Duclert A, Dumas Milne Edwards J, Lacroix B;
PI
XX WPI; 1999-153784/13.
DR
XX N-PSDB; AAX41060.
DR

XX New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from kidney, fetal kidney, dystrophic
PT muscle, muscle and heart tissue
XX
XX Claim 34; Page 601; 622pp; English.
XX
XX AAX40826 to AAX41093 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins, and encode the proteins given in AA101602 and
CC AA111994 to AA12260, respectively. The proteins given represent the

CC signal peptide and an N-terminal fragment of a secreted protein. The
CC nucleic acid sequences can be used for producing secreted human gene
CC products. They can also be used to develop products for diagnosis and
CC therapy. The proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used
CC for directing extracellular secretion of a polypeptide or the insertion
CC of a polypeptide into a membrane, or importing a polypeptide into
CC a cell.

SO Sequence 55 AA:

Query Match 80.6%; Score 29; DB 20; Length 55;
Best Local Similarity 50.0%; Pred. No. 32;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxayxg 8
|||
Db 30 wkssyvg 37

RESULT 12
AAE04286 ID AAE04286 standard; peptide: 56 AA.

AC AAE04286;
XX 09-AUG-2001 (first entry)
DT

DE Human gene 10 encoded secreted protein fragment, SEQ ID NO:150.
XX
XX Human: secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiogenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;
KW endocrine disorder; infection; wound healing; vulnary;
KW cell culture; chemotaxis; food additive;
KW binding partner identification.
XX
XX Homo sapiens.
OS
XX MO200136432-A2.
PN
XX 25-MAY-2001.
PD
XX 15-NOV-2000; 2000MO-US31162.
PE
XX 19-NOV-1999; 99US-0166415.
PR
XX 30-JUN-2000; 2000US-0215136.
PR

XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Komatsoulis GA, Baker KP, Young PE;
PI
XX WPI; 2001-343793/36.
DR

XX Isolated nucleic acid molecule encoding a human secreted protein is
PT used in preventing, treating or ameliorating a medical condition -
PT
XX Disclosure; Page 39; 509pp; English.
XX
XX AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted

CC protein genes, and AAE04199-AAE04239 represent the proteins they encode.
CC AAE04240-AAE04297 represent human secreted protein fragments or variants.
CC The secreted proteins and their genes are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 18 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and immune system
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin ageing due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays (e.g., radioimmunoassay or enzyme linked
CC immunoabsorbent assay (ELISA)). The present sequence represents a human
CC secreted protein fragment referred to in the disclosure of the invention.

XX Sequence 56 AA:

Query Match 80.6%; Score 29; DB 22; Length 56;
Best Local Similarity 50.0%; Pred. No. 33;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxayxg 8
||:|
Db 6 wktsyyvg 13

RESULT 13

AAEA0407
ID AAEA0407 standard; Protein; 83 AA.

XX AAEA0407;

XX AC 22-OCT-2001 (first entry)

XX DT Human polypeptide SEQ ID NO 3552.

XX DE Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;

KW peripheral nervous system; neuropathy; central nervous system; CNS;

KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

KW chemokine; thrombolytic; drug screening; arthritis; inflammation;

KW leukaemia.

XX Homo sapiens.

XX OS WO200153312-A1.

XX PN 26-JUL-2001.

XX PD 26-DEC-2000; 2000WO-US34263.

XX PF 21-JAN-2000; 2000US-0488725.

XX PR 25-APR-2000; 2000US-0552317.

XX PR 19-JUL-2000; 2000US-0598042.

XX PR 03-AUG-2000; 2000US-0620312.

XX PR 14-SEP-2000; 2000US-0653450.

XX PR 19-OCT-2000; 2000US-0662191.

XX PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX MPI: 2001-442253/47.

DR N-PSDB; AAI59563.

XX Novel nucleic acids and polypeptides, useful for treating disorders

PT such as central nervous system injuries -

PS Example 7; SEQ ID NO 3552; 10078pp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and

CC the encoded polypeptides (AAAM8642-AAAM4213) with noctropic,

CC immunosuppressant and cytostatic activity. The polynucleotides are useful

CC in gene therapy. A composition containing a polypeptide or polynucleotide

CC of the invention may be used to treat diseases of the peripheral nervous

CC system, such as peripheral nervous injuries, peripheral neuropathy and

CC localised neuropathies and central nervous system diseases, such as

CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

CC utilisation of the activities such as: Immune system suppression,

CC activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic

CC and thrombolytic activity, cancer diagnosis and therapy, drug screening, and

CC assays for receptor activity, arthritis and inflammation, leukaemias and

CC C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed

CC specification.

XX Sequence 83 AA:

Query Match 80.6%; Score 29; DB 22; Length 83;
Best Local Similarity 50.0%; Pred. No: 48;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxayxg 8
||:|
Db 30 wkqssyag 37

RESULT 14

AAEA04281
ID AAEA04281 standard; Protein; 113 AA.

XX AAEA04281;

XX AC 09-AUG-2001 (first entry)

XX DT Human gene 10 encoded secreted protein fragment, SEQ ID NO:145.

XX DE Human; secreted protein; proliferative disorder; cancer; tumour;

KW foetal abnormality; developmental abnormality; haematopoietic disorder;

KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;

KW inflammation; allergy; neurological disorder; schizophrenia; asthma;

KW Parkinson's disease; cognitive disorder; sepsis; diabetes; atherosclerosis;

KW cardiovascular disorder; angiogenic disorder; kidney disorder;

KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;

KW endocrine disorder; infection; wound healing; vulnerrary;

KW cell culture; chemotaxis; food additive;

XX binding partner identification.

XX Homo sapiens.

XX OS Key

XX Location/Qualifiers

XX Misc-difference 109

XX /label= Unknown

XX /note= "Xaa equals any of the naturally occurring

XX FT

FT L-amino acids*
XX WO200136432-A2.
XX 25-MAY-2001.
XX 15-NOV-2000; 2000WO-US31162.
XX 19-NOV-1999; 99US-0166415.
XX 30-JUN-2000; 2000US-0215136.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, KomatsuJLis GA, Baker KP, Young PE;
XX WPI: 2001-343793/36.
XX Isolated nucleic acid molecule encoding a human secreted protein is
XX used in preventing, treating or ameliorating a medical condition -
XX
XX Disclosure: Page 502; 509pp; English.
XX
XX ADD08486-AMD08529 represent cDNAs corresponding to 18 human secreted
XX protein genes, and AAE04199-AAE04239 represent the proteins they encode.
XX AAE04240-AAE04297 represent human secreted protein fragments or variants.
XX The secreted proteins and their genes are useful for preventing,
XX treating or ameliorating medical conditions, e.g., by protein or gene
XX therapy. Pathological conditions can be diagnosed by determining the
XX amount of the new protein in a sample or by determining the presence of
XX mutations in the new genes. Specific uses are described for each of the
XX 18 genes, based on the tissues in which they are most highly expressed,
XX and include developing products for the diagnosis or treatment of
XX proliferative disorders, cancer, tumours, foetal and developmental
XX abnormalities, haematopoietic disorders, diseases of the immune system,
XX AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
XX allergies, neurological disorders (e.g., Alzheimer's disease,
XX Parkinson's disease), cognitive disorders, schizophrenia, asthma,
XX skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
XX cardiovascular disorders, angogenic disorders, kidney disorders,
XX gastrointestinal disorders, pregnancy-related disorders, endocrine
XX disorders, and infections. The proteins can also be used to aid wound
XX healing and epithelial cell proliferation, to prevent skin ageing due to
XX sunburn, to maintain organs before transplantation, for supporting cell
XX culture of primary tissues, to regenerate tissues, to identify their
XX cognate ligands or binding partners, and in chemotaxis, and can be used
XX as a food additive or preservative to modify storage properties.
XX Antibodies specific for a protein of the invention can be used in
XX alleviating symptoms associated with the disorders mentioned above, and
XX in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
XX immunosorbent assay (ELISA). The present sequence represents a human
XX secreted protein fragment referred to in the disclosure of the invention.
XX
XX Sequence 113 AA:
SQ

DE Human 5' EST secreted protein SEQ ID NO: 539.
XX
XX Human; secreted protein; EST; expressed sequence tag; diagnosis;
XX forensic; gene therapy; chromosome mapping; signal peptide;
XX upstream regulatory sequence; cytokine activity; cell proliferation;
XX differentiation; haematopoiesis regulation; tissue growth regulation;
XX reproductive hormone regulation; chemotactic; chemokine; haemostatic;
XX thrombolytic; anti-inflammatory; tumour inhibition.
XX
XX Homo sapiens.
XX
XX WO9906554-A2.
XX 11-FEB-1999.
XX
XX 31-JUL-1998; 98WO-1B01238.
XX
XX 01-AUG-1997; 97US-0905134.
XX
XX (GEST) GENSET.
XX
XX Duclert A, Dumas Mline Edwards J, Lacroix B;
XX WPI: 1999-153784/13.
XX N-PSDB; AAX41059.
XX
XX New nucleic acids encoding human secreted proteins - obtained from
XX cDNA libraries prepared from kidney, fetal kidney, dystrophic
XX muscle, muscle and heart tissue
XX
XX Claim 34; Page 600-601; 622pp; English.
XX
XX AAX40826 to AAX41093 represent 5' expressed sequence tags (ESTs) for
XX human secreted proteins, and encode the proteins given in AAY01602 and
XX AAY11994 to AAY12260, respectively. The proteins given represent the
XX signal peptide and an N-terminal fragment of a secreted protein. The
XX nucleic acid sequences can be used for producing secreted human gene
XX products. They can also be used to develop products for diagnosis and
XX therapy. The proteins obtained may have cytokine activity, cell
XX proliferation/differentiation activity, haematopoiesis regulating
XX activity, tissue growth regulating activity, reproductive hormone
XX regulating activity, chemotactic/chemokinetic activity, haemostatic and
XX thrombolytic activity, receptor/ligand activity, anti-inflammatory
XX activity, tumour inhibition activity or other activities. The products
XX can be used in forensic, gene therapy and chromosome mapping procedures.
XX The sequences can also be used for obtaining corresponding promoter
XX sequences. The nucleic acids encoding the signal peptide can be used
XX for directing extracellular secretion of a polypeptide or the insertion
XX of a polypeptide into a membrane, or importing a polypeptide into
XX a cell.
XX
XX Sequence 114 AA:
SQ

Query Match 80.6%; Score 29; DB 20; Length 114;
Best Local Similarity 50.0%; Pred. No. 65;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
OY 1 wkxayxg 8
DB 30 wktssyvg 37

RESULT 16
AAM25829
ID AAM25829 standard; Protein; 126 AA.
XX
XX AAM25829;
XX
XX 16-OCT-2001 (first entry)
XX
XX Human protein sequence SEQ ID NO:1344.
XX

KW Human; cancer; HIV infection; human immunodeficiency virus;
 KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
 KW antibacterial; endocrine; cardiant; central nervous system; virucide;
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianemic; anaemia;
 KW antiaggregant; haemostatic; vulnery; antitumor; osteopathic; eczema;
 KW dermatological; antiallergic; antistimatic; antidiabetic; cytostatic;
 KW neuroprotective; antipressant; nootropic; antiparkinsonian; infection;
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 KW cardiac dysfunction; neuropathology; cardiac anapylaxis; autoimmunity;
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 KW neurological disorder.
 OS Homo sapiens.
 XX
 OS WO200153455-A2.
 PN
 XX
 PD 26-JUL-2001.
 PF 22-DEC-2000; 2000WO-US35017.
 PR 23-DEC-1999; 990US-0471275.
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 XX
 PA (HYSE-) HYSEQ INC.
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI: 2001-457603/49.
 DR N-PSDB; AAH99770.
 XX
 PT Isolated human polynucleotides encoding polypeptides, useful for the
 PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
 PT
 XX Claim 20; Page 278; 1217pp; English.
 CC AAH99166 to AAH99904 encode the human proteins given in AM255225 to
 CC AM255963. The proteins can have activities based on the tissues and
 CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
 CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
 CC central nervous system; virucide; anti-HIV; fungicide; antitumor;
 CC cardiovascular; antianemic; antiaggregant; haemostatic; vulnery;
 CC antitumor; osteopathic; dermatological; antiallergic; antistimatic;
 CC antidiabetic; cytostatic; neuroprotective; antipressant; nootropic;
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 CC encoding them can be used in gene therapy, antisense therapy and vaccine
 CC production. The proteins and polynucleotides are useful for screening for
 CC agonists or antagonists of a protein and for the treatment and diagnosis
 CC of disorders associated with the activity of a protein e.g. inflammation,
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 CC neuropathology, cardiac anapylaxis, viral, bacterial, HIV and fungal
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, ulcers,
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
 CC neurological disorders.
 CC
 XX Sequence 126 AA;
 XX

Query Match 80.6%; Score 29; DB 22; Length 126;
 Best Local Similarity 50.0%; Pred. No. 72;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxayxg 8
 Db 73 wkgsyag 80

RESULT 17
 AA42193
 ID AA42193 standard; Protein; 126 AA.
 XX
 XX
 AC AA42193;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 XX Human polypeptide SEQ ID NO 7124.
 DE
 XX
 XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 XX
 OS Homo sapiens.
 OS
 PN WO200153312-A1.
 PD 26-JUL-2001.
 PF 26-DEC-2000; 2000WO-US34263.
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-052317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QH, Zhou P, Goodrich R, Drmanac RT;
 DR
 DR WPI: 2001-442253/47.
 DR N-PSDB; AAI61349.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PT
 XX Example 2; SEQ ID NO 7124; 10078pp; English.
 PS
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AA38642-AA42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localized neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibn activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 CC
 XX Sequence 126 AA;
 XX

Query Match 80.6%; Score 29; DB 22; Length 126;
 Best Local Similarity 50.0%; Pred. No. 72;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxyyg 8
| | |
Db 73 wkqsyvg 80

RESULT 18
AAE04228
ID AAE04228 standard; Protein: 140 AA.

XX AAE04228;
XX
XX 09-AUG-2001 (first entry)

XX Human gene 10 encoded secreted protein HTP1H83, SEQ ID NO: 83.

XX Human: secreted protein; proliferative disorder; cancer; tumour;
XX focal abnormality; developmental abnormality; haematopoietic disorder;
XX immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
XX inflammation; allergy; neurological disorder; Alzheimer's disease;
XX Parkinson's disease; cognitive disorder; schizophrenia; asthma;
XX skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
XX cardiovascular disorder; angioneurotic disorder; kidney disorder;
XX gastrointestinal disorder; pregnancy-related disorder; gene therapy;
XX endocrine disorder; infection; wound healing; vulnerability;
XX cell culture; chemotaxis; food additive;
XX binding partner identification.

XX Homo sapiens.

XX Key Location/Qualifiers
XX Peptide 1..24
XX Protein /label= signal_peptide
XX 25..140
XX /note= "Mature secreted protein"
XX MISC-difference 135
XX /note= "Encoded by GWT"

XX WO200136432-A2.
XX
XX 25-MAY-2001.
XX
XX 15-NOV-2000; 2000MO-US31162.
XX
XX 19-NOV-1999; 99US-0166415.
XX
XX 30-JUN-2000; 2000US-0215136.
XX
XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Komatsoulis GA, Baker KP, Young PE;
XX
XX WPI: 2001-343793/36.
XX
XX N-PSDB; AAD08518.

XX Isolated nucleic acid molecule encoding a human secreted protein is
XX used in preventing, treating or ameliorating a medical condition -
XX
XX Claim 11; Page 459; 509pp; English.

XX AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted
XX protein genes, and AAE04199-AAE04239 represent the proteins they encode.
XX AAE04240-AAE04297 represent human secreted protein fragments or variants.
XX The secreted proteins and their genes are useful for preventing,
XX treating or ameliorating medical conditions, e.g., by protein or gene
XX therapy. Pathological conditions can be diagnosed by determining the
XX amount of the new protein in a sample or by determining the presence of
XX mutations in the new genes. Specific uses are described for each of the
XX 18 genes, based on the tissues in which they are most highly expressed,
XX and include developing products for the diagnosis or treatment of
XX proliferative disorders, cancer, tumours, foetal and developmental
XX abnormalities, haematopoietic disorders, diseases of the immune system,
XX AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
XX allergies, neurological disorders (e.g., Alzheimer's disease,
XX Parkinson's disease), cognitive disorders, schizophrenia, asthma,

CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angioneurotic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin ageing due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein of the invention.

XX
XX
XX Sequence 140 AA:

Query Match 80.6%; Score 29; DB 22; Length 140;
Best Local Similarity 50.0%; Pred. No. 79;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxyyg 8
| | |
Db 30 wktsyvg 37

RESULT 19
AAB54052
ID AAB54052 standard; Protein: 155 AA.

XX AAB54052;
XX
XX 09-MAR-2001 (first entry)

XX Human pancreatic cancer antigen protein sequence SEQ ID NO: 504.

XX
XX Human: pancreas; pancreatic cancer; pancreatic cancer antigen;
XX detection; diagnosis; identification; cytostatic; neuroprotective;
XX neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological;
XX antiinflammatory; cardiant; gene therapy; chromosome mapping;
XX linkage analysis; tissue identification; tissue typing; forensic;
XX neutral; immune system; muscular; reproductive; gastrointestinal;
XX pulmonary; cardiovascular; renal; proliferative.

XX Homo sapiens.

XX WO200055320-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000MO-US05989.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;
XX
XX WPI: 2000-579444/54.
XX
XX N-PSDB; AAC98817.

XX New nucleic acid that is a pancreatic cancer antigen for preventing,
XX treating, or ameliorating a medical condition, particular pancreatic
XX cancer, or for use in assays for diagnosing a pathological condition -
XX
XX Claim 11; Page 942-943; 1379pp; English.

XX AAC98773 to AAC99231 encode the human pancreatic cancer associated
XX proteins, called pancreatic cancer antigens, given in AAB54008 to
XX AAB54466. The human pancreatic cancer antigens have cytostatic,
XX neuroprotective, neoplastic, immunomodulatory, relaxant, contraceptive,
XX gynaecological, cardiant and antiinflammatory activities, and can be used

in gene therapy. The polynucleotide and proteins can be used for preventing, treating, or ameliorating a medical condition or in assays for diagnosing a pathological condition or a susceptibility to one in a subject. Binding partners to the proteins and the activity of the proteins can be identified. The pancreatic cancer antigens can be used to detect, treat or prevent pancreatic disorders, especially cancer. Agonists and antagonists to the antigens can be screened for. The pancreatic cancer antigen polynucleotides can be used to design nucleic acid hybridisation probes that can be used in chromosome mapping, linkage analysis, tissue identification and/or typing and a variety of forensic and diagnostic methods. The proteins can be used to generate antibodies which are used to purify, detect and target the polypeptides, including both in vivo and in vitro diagnostic and therapeutic methods. The proteins can be used to treat or prevent neural, immune system, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal or proliferative disorders. AAC99232 to AAC99240 and AAB34467 represent sequences used in the exemplification of the present invention.

Sequence 155 AA;

Query Match 80.6%; Score 29; DB 21; Length 155;
Best Local Similarity 50.0%; Pred. No. 87;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxayxg 8
|| : ||
Db 55 wktssyvg 62

RESULT 20

AAV36134
ID AAV36134 standard; protein; 230 AA.

XX AAV36134;

XX 23-SEP-1999 (first entry)

XX Human secreted protein #6.

XX Secreted protein; human; cytostatic; thrombotic; osteopathic; forensic;

KW diagnostic; gene therapy; chromosome mapping; secretion vector.

XX Homo sapiens.

XX W09925825-A2.

XX 27-MAY-1999.

XX 13-NOV-1998; 98WO-IB01862.

XX 04-SEP-1998; 98US-0099273.

XX 13-NOV-1997; 97US-0066677.

XX 17-DEC-1997; 97US-0069957.

XX 09-FEB-1998; 98US-0074121.

XX 13-APR-1998; 98US-0081563.

XX 10-AUG-1998; 98US-0096116.

XX (GEST) GENSET.

XX Bougueleret L, Duclert A, Dumas Milne Edwards J;

XX WPI, 1999-347472/29.

XX N-PSDB; AAX97818.

XX Extended cDNAs encoding secreted proteins

XX Example 28; Page 234-235; 307pp; English.

XX AAY36129-Y36222 represent novel human secreted proteins encoded by the

XX extended cDNA sequences represented in AAX97813-X97906. The proteins

XX of the invention have cytosolic, thrombotic and osteopathic activity.

XX The extended cDNAs can be used to express secreted proteins or parts of

CC them or to obtain antibodies capable of binding to the secreted

CC proteins. They may also be used in diagnostic, forensic, gene therapy

CC and chromosome mapping procedures. Uses also include design of

CC expression vectors and secretion vectors.

XX

SQ Sequence 230 AA;

Query Match 80.6%; Score 29; DB 20; Length 230;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxayxg 8
|| : ||
Db 30 wktssyvg 37

Search completed: January 14, 2002, 07:35:59
Job time: 810 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:23:41 ; Search time 48.05 Seconds
(without alignments)
3.747 Million cell updates/sec

Title: 09-185908-1b
Perfect score: 36
Sequence: 1 wxxxxxxg 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 segs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCITUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	75.0	376	1	US-08-608-241-2
2	27	75.0	376	2	US-08-922-182-2
3	27	75.0	376	2	US-08-919-953-2
4	27	75.0	376	4	US-09-192-983-2
5	26	72.2	374	2	US-07-857-224B-84
6	26	72.2	759	2	US-08-637-759B-89
7	26	72.2	759	3	US-08-871-355A-89
8	26	72.2	2987	2	US-08-970-269A-29
9	26	72.2	2987	4	US-09-407-562-29
10	26	72.2	3959	4	US-08-970-269A-30
11	26	72.2	3959	4	US-09-407-562-30
12	25	69.4	63	2	US-08-993-228-14
13	25	69.4	235	1	US-07-971-160-4
14	25	69.4	235	1	US-08-336-241-4
15	25	69.4	235	2	US-08-465-273-4
16	25	69.4	235	2	US-09-119-024-4
17	25	69.4	235	2	US-08-417-226-4
18	25	69.4	235	4	US-09-196-131-4
19	25	69.4	237	1	US-07-971-160-2
20	25	69.4	237	1	US-07-971-160-2
21	25	69.4	237	1	US-08-336-241-2
22	25	69.4	237	1	US-08-336-241-16
23	25	69.4	237	2	US-08-465-273-2
24	25	69.4	237	2	US-08-465-273-16
25	25	69.4	237	2	US-09-119-024-2
26	25	69.4	237	2	US-09-119-024-16
27	25	69.4	237	2	US-08-417-226-2

28	25	69.4	237	2	US-08-417-226-16	Sequence 16, Appl
29	25	69.4	237	4	US-09-196-131-2	Sequence 2, Appl
30	25	69.4	237	4	US-09-196-131-16	Sequence 16, Appl
31	25	69.4	344	2	US-07-857-224B-92	Sequence 92, Appl
32	25	69.4	344	2	US-07-857-224B-93	Sequence 93, Appl
33	25	69.4	432	2	US-08-677-049-8	Sequence 8, Appl
34	25	69.4	464	2	US-09-021-323-3	Sequence 3, Appl
35	25	69.4	554	1	US-08-445-586-7	Sequence 7, Appl
36	25	69.4	556	1	US-08-445-586-2	Sequence 2, Appl
37	25	69.4	645	1	US-07-779-172A-3	Sequence 3, Appl
38	25	69.4	1620	1	US-08-542-363-2	Sequence 2, Appl
39	25	69.4	1620	4	US-09-100-089-2	Sequence 2, Appl
40	25	69.4	1663	2	US-08-793-126-1	Sequence 1, Appl
41	25	69.4	1663	4	US-09-132-271-1	Sequence 1, Appl
42	25	69.4	1663	4	US-09-142-334-22	Sequence 22, Appl
43	24	66.7	222	4	US-09-140-804-7	Sequence 7, Appl
44	24	66.7	345	3	US-09-222-817-2	Sequence 2, Appl
45	24	66.7	345	4	US-09-222-817-2	Sequence 2, Appl
46	24	66.7	422	2	US-08-484-575A-6	Sequence 6, Appl
47	24	66.7	422	3	US-08-477-459-6	Sequence 6, Appl
48	24	66.7	422	3	US-08-479-869-6	Sequence 6, Appl
49	24	66.7	422	4	US-08-486-414-6	Sequence 6, Appl
50	24	66.7	422	5	PCT-US94-01826A-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-08-608-241-2
Sequence 2, Application US/08608241
Patent No. 5747328
GENERAL INFORMATION:
APPLICANT: Donohue, Timothy J
APPLICANT: Barber, Robert D
APPLICANT: Withuhn, Vernon
TITLE OF INVENTION: MICROBIAL SYSTEM FOR FORMALDEHYDE
NUMBER OF SEQUENCES: 7
TITLE OF INVENTION: SENSING AND REMEDIATION
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Plinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/608,241
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J
REGISTRATION NUMBER: 27,386
REFERENCE/DOCKET NUMBER: 960296,93511
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 376 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-608-241-2

Query Match 75.0% Score 27; DB 1; Length 376;
Best Local Similarity 50.0%; Pred. No. 1.9e+02;

Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxayxg 8
|| | : |
Db 316 WKGSAFEG 323

RESULT 2
US-08-922-182-2
; Sequence 2, Application US/08922182
; Patent No. 5834300
; GENERAL INFORMATION:
; APPLICANT: Donohue, Timothy J
; APPLICANT: Barber, Robert D
; APPLICANT: Withuhn, Vernon
; TITLE OF INVENTION: MICROBIAL SYSTEM FOR FORMALDEHYDE
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/922,182
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/608,241
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 960296.93511
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 376 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-922-182-2

Query Match 75.0%; Score 27; DB 2; Length 376;
Best Local Similarity 50.0%; Pred. No. 1.9e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxayxg 8
|| | : |
Db 316 WKGSAFEG 323

RESULT 3
US-08-919-953-2
; Sequence 2, Application US/08919953
; Patent No. 5837481
; GENERAL INFORMATION:
; APPLICANT: Donohue, Timothy J
; APPLICANT: Barber, Robert D
; APPLICANT: Withuhn, Vernon
; TITLE OF INVENTION: MICROBIAL SYSTEM FOR FORMALDEHYDE
; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/919,953
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/608,241
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 960296.93511
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 376 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-919-953-2

Query Match 75.0%; Score 27; DB 2; Length 376;
Best Local Similarity 50.0%; Pred. No. 1.9e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxayxg 8
|| | : |
Db 316 WKGSAFEG 323

RESULT 4
US-09-192-983-2
; Sequence 2, Application US/09192983A
; Patent No. 6242244
; GENERAL INFORMATION:
; APPLICANT: Donohue, Timothy
; APPLICANT: Barber, Robert
; APPLICANT: Withuhn, Vernon
; TITLE OF INVENTION: Microbial System for Formaldehyde Sensing and
; FILE REFERENCE: 960296.95505
; CURRENT APPLICATION NUMBER: US/09/192,983A
; CURRENT FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: 08/919,953
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 08/608,241
; EARLIER FILING DATE: 1996-02-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Rhodobacter sphaeroides
US-09-192-983-2

Query Match 75.0%; Score 27; DB 4; Length 376;
Best Local Similarity 50.0%; Pred. No. 1.9e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
OY      1 wxxxxxg 8
      || | | |
Db      316 WKGSAFCG 323

RESULT  5
US-07-857-224B-84
; Sequence 84, Application US/07857224B
; Patent No. 5958784
; GENERAL INFORMATION:
; APPLICANT: Benner, Steven A.
; TITLE OF INVENTION: Predicting Folded Structures of Proteins
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven A. Benner
; STREET: Hadlaubstrasse 151
; CITY: Zurich
; STATE: none
; COUNTRY: Switzerland
; ZIP: (note: this is an International post code) CH-8092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/857,224B
; FILING DATE: 03/25/92
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA: none
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (International) 41 1 632 2830
; TELEFAX: (International) 41 1 262 2437
; TELEX: none
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 374
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: protein
; ORIGINAL SOURCE:
; ORGANISM: human
; FEATURE: Alcohol dehydrogenase, Table 3 Column 5
; PUBLICATION INFORMATION:
; AUTHORS:
; AUTHORS: Joernvall, H.
; AUTHORS: Persson, M.
; AUTHORS: Jeffery, J.
; TITLE: Alcohol dehydrogenases
; JOURNAL: Proceedings of the National Academy of Sciences, USA
; VOLUME: 78
; PAGES: 4226-4230
; DATE: 1981
;
US-07-857-224B-84

Query Match      72.2%: Score 26; DB 2; Length 374;
Best Local Similarity 50.0%: Pred. No. 3e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
OY      1 wxxxxxg 8
      || | | |
Db      314 WKGATYGG 321
```

```
RESULT  6
US-08-637-759B-89
; Sequence 89, Application US/08637759B
; Patent No. 5876931
; GENERAL INFORMATION:
; APPLICANT: David William Holden
```

```
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/637,759B
; FILING DATE: 03-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/02875
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPKS 101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 759 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
;
US-08-637-759B-89
```

```
Query Match      72.2%: Score 26; DB 2; Length 759;
Best Local Similarity 50.0%: Pred. No. 5.6e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
OY      1 wxxxxxg 8
      || | | |
Db      30 WKPAAFQG 37
```

```
RESULT  7
US-08-871-355A-89
; Sequence 89, Application US/08871355A
; Patent No. 6013669
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/871,355A
```

```

; FILING DATE: 09-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/02875
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 759 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-08-871-355A-89

```

```

Query Match          72.2%; Score 26; DB 3; Length 759;
Best Local Similarity 50.0%; Pred. No. 5.6e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

```

OY 1 wkxxayxg 8
    11 11
Db 30 WKPAAFQG 37

```

```

RESULT 8
US-08-970-269A-29
; Sequence 29, Application US/08970269A
; Patent No. 5976803
; GENERAL INFORMATION:
; APPLICANT: Kathryn Meek
; TITLE OF INVENTION: Genetic Test For Equine Severe
; TITLE OF INVENTION: Combined Immunodeficiency Disease
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Dr. Benjamin A. Adler
; STREET: 8011 Candle Lane
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word for Macintosh
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/970,269A
; FILING DATE: No. 5976803ember 14, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Adler Ph.D., Benjamin A.
; REGISTRATION NUMBER: 35,423
; REFERENCE/DOCKET NUMBER: D5860
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-777-2321
; TELEFAX: 713-777-6908
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2987 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: no

```

```

; ANTI-SENSE: no
; ORIGINAL SOURCE:
; FEATURE:
; US-08-970-269A-29

```

```

Query Match          72.2%; Score 26; DB 2; Length 2987;
Best Local Similarity 66.7%; Pred. No. 1.9e+03;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 1 wkxxay 6
    11 11
Db 2840 WKSLEY 2845

```

```

RESULT 9
US-09-407-562-29
; Sequence 29, Application US/09407562
; Patent No. 6294334
; GENERAL INFORMATION:

```

```

; APPLICANT: Kathryn Meek
; TITLE OF INVENTION: Genetic Test For Equine Severe
; TITLE OF INVENTION: Combined Immunodeficiency Disease
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Dr. Benjamin A. Adler
; STREET: 8011 Candle Lane
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word for Macintosh
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/407,562
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/970,269
; FILING DATE: No. 6294334ember 14, 1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Adler Ph.D., Benjamin A.
; REGISTRATION NUMBER: 35,423
; REFERENCE/DOCKET NUMBER: D5860
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-777-2321
; TELEFAX: 713-777-6908
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2987 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: no
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; FEATURE:
; US-09-407-562-29

```

```

Query Match          72.2%; Score 26; DB 4; Length 2987;
Best Local Similarity 66.7%; Pred. No. 1.9e+03;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 1 wkxxay 6
    11 11
Db 2840 WKSLEY 2845

```

```
RESULT 10
US-08-970-269A-30
: Sequence 30, Application US/08970269A
: Patent No. 5976803
: GENERAL INFORMATION:
: APPLICANT: Katherine Meek
: TITLE OF INVENTION: Genetic Test For Equine Severe
: NUMBER OF SEQUENCES: 32
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dr. Benjamin A. Adler
: STREET: 8011 Candle Lane
: CITY: Houston
: STATE: Texas
: COUNTRY: USA
: ZIP: 77071
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: Apple
: OPERATING SYSTEM: Macintosh
: SOFTWARE: Microsoft Word for Macintosh
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/970,269A
: FILING DATE: No. 5976803ember 14, 1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Adler Ph.D., Benjamin A.
: REGISTRATION NUMBER: 35,423
: REFERENCE/DOCKET NUMBER: D5860
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 713-777-2321
: TELEFAX: 713-777-6908
: INFORMATION FOR SEQ ID NO: 30:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3959 amino acid
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHETICAL: no
: ANTI-SENSE: no
: US-08-970-269A-30

Query Match 72.2%; Score 26; DB 2; Length 3959;
Best Local Similarity 66.7%; Pred. No. 2.5e+03;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 Wkxay 6
DB 2840 Wkslay 2845

RESULT 11
US-09-407-562-30
: Sequence 30, Application US/09407562
: Patent No. 6294334
: GENERAL INFORMATION:
: APPLICANT: Katherine Meek
: TITLE OF INVENTION: Genetic Test For Equine Severe
: NUMBER OF SEQUENCES: 32
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dr. Benjamin A. Adler
: STREET: 8011 Candle Lane
: CITY: Houston
: STATE: Texas
: COUNTRY: USA
: ZIP: 77071
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: Apple
```

```
OPERATING SYSTEM: Macintosh
SOFTWARE: Microsoft Word for Macintosh
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/407,562
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/970,269
FILING DATE: No. 6294334ember 14, 1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Adler Ph.D., Benjamin A.
REGISTRATION NUMBER: 35,423
REFERENCE/DOCKET NUMBER: D5860
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-777-2321
TELEFAX: 713-777-6908
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 3959 amino acid
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: no
ANTI-SENSE: no
US-09-407-562-30

Query Match 72.2%; Score 26; DB 4; Length 3959;
Best Local Similarity 66.7%; Pred. No. 2.5e+03;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 Wkxay 6
DB 2840 Wkslay 2845

RESULT 12
US-08-993-228-14
: Sequence 14, Application US/08993228
: Patent No. 5976838
: GENERAL INFORMATION:
: APPLICANT: Jacobs, Kenneth
: APPLICANT: McCoy, John M.
: APPLICANT: Lavallie, Edward R.
: APPLICANT: Racie, Lisa A.
: APPLICANT: Merberg, David
: APPLICANT: Treacy, Maurice
: APPLICANT: Spaulding, Vikki
: APPLICANT: Agostino, Michael J.
: TITLE OF INVENTION: SECRETED PROPEINS AND POLYNUCLEOTIDES
: NUMBER OF SEQUENCES: 33
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genetics Institute, Inc.
: STREET: 87 Cambridgepark Drive
: CITY: Cambridge
: STATE: MA
: COUNTRY: U.S.A.
: ZIP: 02140
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/993,228
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Sprunger, Suzanne A.
: REGISTRATION NUMBER: 41,323
```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 63 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-993-228-14

Query Match 69.4%; Score 25; DB 2; Length 63;
Best Local Similarity 50.0%; Pred. No. 95;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wxxxayxg 8
|| ||
Db 51 WKNKAYRG 58

RESULT 13
US-07-971-160-4
; Sequence 4, Application US/07971160
; Patent No. 5474896
; GENERAL INFORMATION:
; APPLICANT: Dujon, Bernard
; APPLICANT: Chouluka, Andre
; APPLICANT: Colleaux, Laurence
; APPLICANT: Fairhead, Cecile
; APPLICANT: Perrin, Arnaud
; APPLICANT: Plessis, Anne
; APPLICANT: Thierry, Agnes
; TITLE OF INVENTION: Nucleotide Sequence Encoding the Enzyme
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farbow, Garrett &
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/971,160
; FILING DATE: 19921105
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/879,689
; FILING DATE: 05-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Ogden, Stasia L.
; REGISTRATION NUMBER: 36,228
; REFERENCE/DOCKET NUMBER: 03495-0111-01000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 235 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-971-160-4

Query Match 69.4%; Score 25; DB 1; Length 235;
Best Local Similarity 66.7%; Pred. No. 3,1e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 wxxxay 6
|| ||
Db 62 WKNKAY 67

RESULT 14
US-08-336-241-4
; Sequence 4, Application US/08336241
; Patent No. 5792632
; GENERAL INFORMATION:
; APPLICANT: Chouluka, Andre
; APPLICANT: Perrin, Arnaud
; APPLICANT: Dujon, Bernard
; APPLICANT: Nicolas, Jean-Francois
; TITLE OF INVENTION: Nucleotide Sequence Encoding the Enzyme
; Patent No. 5792632
; TITLE OF INVENTION: I-SCFI and the Uses Thereof
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farbow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,241
; FILING DATE: 07-NOV-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/971,160
; FILING DATE: 05-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/879,689
; FILING DATE: 05-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E.R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 03495-0111-03000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 235 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-336-241-4

Query Match 69.4%; Score 25; DB 1; Length 235;
Best Local Similarity 66.7%; Pred. No. 3,1e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 wxxxay 6
|| ||
Db 62 WKNKAY 67

RESULT 15
US-08-465-273-4
; Sequence 4, Application US/08465273

```
Patent No. 5866361
GENERAL INFORMATION:
APPLICANT: Choulika, Andre
APPLICANT: Perrin, Arnaud
APPLICANT: Dujon, Bernard
APPLICANT: Nicolas, Jean-Francois
TITLE OF INVENTION: Nucleotide Sequence Encoding the Enzyme
Patent No. 5866361
TITLE OF INVENTION: I-SCI and the Uses Thereof
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,273
FILING DATE: 06-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/336,241
FILING DATE: 07-NOV-1994
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,160
FILING DATE: 05-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/879,689
FILING DATE: 05-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Potler, Jane E.R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 03495-0111-06000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4000
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 235 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-465-273-4

Query Match          69.4%; Score 25; DB 2; Length 235;
Best Local Similarity 66.7%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 wxxxay 6
   1111
Db 62 WKNKAY 67

RESULT 16
US-09-119-024-4
; Sequence 4, Application US/09119024
; GENERAL INFORMATION:
; APPLICANT: Choulika, Andre
; APPLICANT: Perrin, Arnaud
; APPLICANT: Dujon, Bernard
; APPLICANT: Nicolas, Jean-Francois
; TITLE OF INVENTION: Nucleotide Sequence Encoding the Enzyme
; Patent No. 5948678
```

```
TITLE OF INVENTION: I-SCI and the Uses Thereof
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/119,024
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/336,241
FILING DATE: 07-NOV-1994
APPLICATION NUMBER: US 07/971,160
FILING DATE: 05-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/879,689
FILING DATE: 05-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Potler, Jane E.R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 03495-0111-03000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4000
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 235 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-119-024-4

Query Match          69.4%; Score 25; DB 2; Length 235;
Best Local Similarity 66.7%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 wxxxay 6
   1111
Db 62 WKNKAY 67

RESULT 17
US-08-417-226-4
; Sequence 4, Application US/08417226
; Patent No. 5962327
; GENERAL INFORMATION:
; APPLICANT: Choulika, Andre
; APPLICANT: Perrin, Arnaud
; APPLICANT: Dujon, Bernard
; APPLICANT: Nicolas, Jean-Francois
; TITLE OF INVENTION: Nucleotide Sequence Encoding the Enzyme
; Patent No. 5962327
; TITLE OF INVENTION: I-SCI and the Uses Thereof
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/417,226
FILING DATE: 05-APRIL-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/336,241
FILING DATE: 07-NOV-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/971,160
FILING DATE: 05-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/879,689
FILING DATE: 05-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Potter, Jane E.R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 03495-0111-04000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4400
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 235 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-417-226-4

Query Match 69.4%; Score 25; DB 2; Length 235;
Best Local Similarity 66.7%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 wxxxay 6
DB 62 WKNKAY 67

RESULT 18
US-09-196-131-4
Sequence 4, Application US/09196131
Patent No. 6238924
GENERAL INFORMATION:
APPLICANT: Chouluka, Andre
APPLICANT: Perrin, Arnaud
APPLICANT: Dujon, Bernard
APPLICANT: Nicolas, Jean-Francois
TITLE OF INVENTION: Nucleotide Sequence Encoding the Enzyme
Patent No. 6238924
TITLE OF INVENTION: I-SCEI and the Uses Thereof
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farbow, Garrett &
ADDRESS: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/196,131
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/417,226
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/879,689
FILING DATE: 05-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Potter, Jane E.R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 03495-0111-04000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4400
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 235 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-196-131-4

Query Match 69.4%; Score 25; DB 4; Length 235;
Best Local Similarity 66.7%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 wxxxay 6
DB 62 WKNKAY 67

RESULT 19
US-07-971-160-2
Sequence 2, Application US/07971160
Patent No. 5474896
GENERAL INFORMATION:
APPLICANT: Dujon, Bernard
APPLICANT: Chouluka, Andre
APPLICANT: Colleaux, Laurence
APPLICANT: Falthead, Cecile
APPLICANT: Plessis, Anne
APPLICANT: Thierry, Agnes
TITLE OF INVENTION: Nucleotide Sequence Encoding the Enzyme
TITLE OF INVENTION: I-SCEI and the Uses Thereof
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farbow, Garrett &
ADDRESS: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/971,160
FILING DATE: 19921105
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/879,689
FILING DATE: 05-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Ogden, Stasia L.
REGISTRATION NUMBER: 36,228
REFERENCE/DOCKET NUMBER: 03495-0111-01000
TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 237 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-07-971-160-2

Query Match 69.4%; Score 25; DB 1; Length 237;
Best Local Similarity 66.7%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 wKxxay 6
DB 64 WKNKAY 69

RESULT 20

US-07-971-160-16
Sequence 16, Application US/07971160
Patent No. 5474896

GENERAL INFORMATION:

APPLICANT: DuJon, Bernard
APPLICANT: Choulka, Andre
APPLICANT: Colleaux, Laurence
APPLICANT: Fairhead, Cecile
APPLICANT: Perrin, Arnaud
APPLICANT: Plesels, Anne
APPLICANT: Thierry, Agnes
TITLE OF INVENTION: Nucleotide Sequence Encoding the Enzyme
TITLE OF INVENTION: I-SCEI and the Uses Thereof
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finegan, Henderson, Farbow, Garrett &
ADDRESSER: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/971,160
FILING DATE: 19921105
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/879,689
FILING DATE: 05-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Ogden, Stasla L.
REGISTRATION NUMBER: 36,228
REFERENCE/DOCKET NUMBER: 03495-0111-01000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 237 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-07-971-160-16

Query Match 69.4%; Score 25; DB 1; Length 237;
Best Local Similarity 66.7%; Pred. No. 3.1e+02;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 wKxxay 6
DB 64 WKNKAY 69

Search completed: January 14, 2002, 07:23:42
Job time: 73 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:37:26 ; Search time 63.57 Seconds
(without alignments)
9.586 Million cell updates/sec

Title: 09-185908-1c
Perfect score: 35
Sequence: 1 wkxsfxg 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	88.6	528	2	G70854
2	31	88.6	528	2	T45418
3	29	82.9	392	2	D70475
4	29	82.9	2347	1	TVHURS
5	28	80.0	193	2	A37873
6	28	80.0	206	2	H75258
7	28	80.0	213	2	A56152
8	28	80.0	224	2	A60032
9	28	80.0	428	2	C96633
10	28	80.0	475	2	S65290
11	28	80.0	599	2	A29947
12	28	80.0	600	2	A28960
13	28	80.0	600	2	S00561
14	27	77.1	65	2	S23164
15	27	77.1	369	1	D64763
16	27	77.1	369	1	S57525
17	27	77.1	369	1	C85530
18	27	77.1	373	1	A33419
19	27	77.1	373	1	DEIRA
20	27	77.1	373	1	S68061
21	27	77.1	374	1	DEHUC2
22	27	77.1	374	1	A56643
23	27	77.1	376	1	JC4967
24	27	77.1	376	1	SS1187
25	27	77.1	378	1	A49662
26	27	77.1	378	1	H64052
27	27	77.1	378	2	F81097
28	27	77.1	379	1	SS1357
29	27	77.1	379	1	S71244

30	27	77.1	379	2	D81515	cell shape-determi
31	27	77.1	379	2	H81741	cell shape-determi
32	27	77.1	379	2	A71479	probable rod shape
33	27	77.1	381	1	JN0447	alcohol dehydrogen
34	27	77.1	381	2	T03289	formaldehyde dehyd
35	27	77.1	381	2	T04164	formaldehyde dehyd
36	27	77.1	386	1	S31140	alcohol dehydrogen
37	27	77.1	396	2	S31959	acyl-l-acyl-carrier
38	27	77.1	415	2	A86599	rod shape protein
39	27	77.1	415	2	C72026	rod shape protein
40	27	77.1	441	2	B84854	hypothetical prote
41	27	77.1	599	2	T18316	hypothetical prote
42	27	77.1	733	1	S33643	transforming prote
43	27	77.1	1083	2	S54293	regulator protein
44	27	77.1	1956	2	T16416	hypothetical prote
45	26	74.3	135	2	J01225	unc1 protein homol
46	26	74.3	159	1	OCQV2	apid transmissio
47	26	74.3	199	1	TUBPKL	tail assembly prot
48	26	74.3	200	2	T02779	y4dp protein - Rhl
49	26	74.3	224	2	B85584	probable tail comp
50	26	74.3	246	2	T32510	hypothetical prote

ALIGNMENTS

RESULT 1
G70854
probable sera protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: G70854
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
R:Coile, S.T.; Davies, R.; Devlin, K.; Fellwell, T.; Genies, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Sgares, R.; Stuston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: G70854
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-528 <COL>
A:Cross-references: GB:AL021287; GB:AL123456; NID:93261508; PIDN:CAA16081.1; PID:9279
C:Experimental source: strain H37RV
C:Genetics:
A:Gene: sera
C:Superfamily: Bacillus phosphoglycerate dehydrogenase
Query Match 88.6%; Score 31; DB 2; Length 528;
Best Local Similarity 62.5%; Pred. No. 22;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 wkxsfxg 8
DB 129 WKRSSFSG 136
RESULT 2
T45418
phosphoglycerate dehydrogenase [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C>Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 18-Feb-2000
C:Accession: T45418
R:Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1997
A:Reference number: T45418
A:Accession: T45418
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-528 <PAR>
A:Cross-references: EMBL:Z99263; PIDN:CAB16440.1

A:Experimental source: cosmid B637
C:Genetics:
A:Note: sera
C:Superfamily: Bacillus phosphoglycerate dehydrogenase

Query Match 88.6%; Score 31; DB 2; Length 528;
Best Local Similarity 62.5%; Pred. No. 22;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxsfxg 8
|||
Db 129 WKRSFSG 136

RESULT 3
D70475
conserved hypothetical protein aq_2044 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 17-Mar-2000
C:Accession: D70475
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666
A:Accession: D70475
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-392 <AOF>
A:Cross-references: GB:AE000769; NID:g2984262; PIDN:AAC07799.1; PID:g2984273; GB:AE00065
C:Genetics:
A:Gene: aq_2044
C:Superfamily: comb protein

Query Match 82.9%; Score 29; DB 2; Length 392;
Best Local Similarity 62.5%; Pred. No. 46;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxsfxg 8
|||
Db 127 WKRSFSG 134

RESULT 4
TYHURS
kinase-related protein ros-1 precursor - human
N:Alternate names: protein-tyrosine kinase mcfs (activated ros-1)
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) ros-1
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1988 #sequence_revision 07-Oct-1994 #text_change 11-Jun-1999
C:Accession: A35512; A25223; A24421; A33081
R:Birchmeier, C.; O'Neill, K.; Riggs, M.; Wigler, M.
Proc. Natl. Acad. Sci. U.S.A. 87, 4799-4803, 1990
A:Title: Characterization of ROS1 cDNA from a human glioblastoma cell line.
A:Reference number: A35512; MUID:90280463
A:Accession: A35512
A:Molecule type: mRNA
A:Residues: 1-2212, 'N', 2214-2227, 'QC', 2229-2347 <BIR>
A:Cross-references: GB:M34353
R:Experimental source: glioblastoma cell line SW-1088
R:Matsumine, H.; Wang, L.H.; Shibuya, M.
Mol. Cell. Biol. 6, 3000-3004, 1986
A:Title: Human c-ros-1 gene homologous to the v-ros sequence of UR2 sarcoma virus encode
A:Reference number: A25223; MUID:87064611
A:Accession: A25223
A:Molecule type: DNA
A:Residues: 1790-2245, 'KEDSSSEFSFRCTVN' <MA2>
A:Cross-references: GB:M13368
A:Experimental source: placenta
A:Note: the differences after residue 2245 result from the authors' misinterpretation of

R:Birchmeier, C.; Birnbaum, D.; Matches, G.; Fasano, O.; Wigler, M.
Mol. Cell. Biol. 6, 3109-3116, 1986
A:Title: Characterization of an activated human ros gene.
A:Reference number: A24421; MUID:87064625
A:Accession: A24421
A:Molecule type: mRNA
A:Residues: 1854-2261, 'A', 2263-2347 <B12>
A:Cross-references: GB:M13880; NID:g337482; PIDN:AAA36580.1; PID:g337483
A:Experimental source: tumor cells
A:Note: The mcfs oncogene was formed by DNA rearrangement involving fusion of at least
C:Genetics:
A:Gene: GDB:ROS1
A:Cross-references: GDB:120351; OMIM:165020
A:Map position: 6q22-6q22
A:Introns: 1853/1; 1881/1; 1926/2; 1980/3; 2002/2; 2045/3; 2078/2; 2145/2; 2190/2
C:Superfamily: Kinase-related protein ros; LDL receptor YWTD-containing repeat homolo
C:Keywords: ATP; autophosphorylation; glycoprotein; Kinase-related transforming prote
F:1-36/Domain: signal sequence #status predicted <SIG>
F:37-2347/Product: kinase-related protein ROS1 #status predicted <MAT>
F:37-1859/Domain: extracellular #status predicted <EXT>
F:335-378/Domain: LDL receptor YWTD-containing repeat homology <YW1>
F:466-503/Domain: LDL receptor YWTD-containing repeat homology <YW2>
F:715-757/Domain: LDL receptor YWTD-containing repeat homology <YW3>
F:758-798/Domain: LDL receptor YWTD-containing repeat homology <YW4>
F:799-838/Domain: LDL receptor YWTD-containing repeat homology <YW5>
F:843-888/Domain: LDL receptor YWTD-containing repeat homology <YW6>
F:893-933/Domain: LDL receptor YWTD-containing repeat homology <YW7>
F:1532-1574/Domain: LDL receptor YWTD-containing repeat homology <YW8>
F:1860-1883/Domain: transmembrane #status predicted <INT>
F:1884-2347/Domain: intracellular #status predicted <INT>
F:1943-2222/Domain: protein kinase homology <KIN>
F:1951-1959/Region: protein kinase ATP-binding motif
F:52,114,123,324,352,471,607,628,706,714,732,939,961,1015,1087,1090,1211,1272,1330,14
F:1980/Active site: lys #status predicted
F:2110,2114,2115/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #s

Query Match 82.9%; Score 29; DB 1; Length 2347;
Best Local Similarity 50.0%; Pred. No. 2,6e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxsfxg 8
|||
Db 119 WKSANFSG 126

RESULT 5
A37873
cerebellin precursor - human
C:Species: Homo sapiens (man)
C:Date: 30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change 21-Jul-2000
C:Accession: A37873
R:Urade, Y.; Oberdick, J.; Molinar-Rode, R.; Morgan, J.I.
Proc. Natl. Acad. Sci. U.S.A. 88, 1069-1073, 1991
A:Title: Precerebellin is a cerebellum-specific protein with similarity to the globul
A:Reference number: A37873; MUID:91126057
A:Accession: A37873
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-193 <URA>
A:Cross-references: GB:M58583; NID:g180250; PIDN:AAA3676.1; PID:g180251
C:Superfamily: complement subcomponent C1q chain A; complement C1q carboxyl-terminal
F:63-192/Domain: complement C1q carboxyl-terminal homology <C1q>

Query Match 80.0%; Score 28; DB 2; Length 193;
Best Local Similarity 50.0%; Pred. No. 39;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxsfxg 8
|||
Db 180 WKYSFSG 187

```
RESULT 6
H75258
Probable 3-demethylubiquinone-9-3-methyltransferase - Deinococcus radiodurans (strain R1
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: H75258
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036696
A:Accession: H75258
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-206 <WHI>
A:Cross-references: GB:AE002085; GB:AE000513; NID:g6460383; PIDN:AAF12100.1; PID:g646038
C:Genetics:
A:Gene: DR2562
A:Map position: 1

Query Match
Best Local Similarity 80.0%; Score 28; DB 2; Length 206;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsfxg 8
| | | |
Db 111 WKRAFSPG 118

RESULT 7
A56152
Major 25k outer membrane protein precursor - Brucella abortus
C:Species: Brucella abortus
C:Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 08-Oct-1999
C:Accession: A56152
R:de Wergifosse, P.; Linhermans, P.; Ijmet, J.N.; Cloeckaert, A.
J. Bacteriol. 177, 1911-1914, 1995
A:Title: Cloning and nucleotide sequence of the gene coding for the major 25-kilodalton
A:Reference number: A56152; MUID:95204367
A:Accession: A56152
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-213 <DEA>
A:Cross-references: GB:X79284; NID:g769744; PIDN:CAA55872.1; PID:g769745

Query Match
Best Local Similarity 80.0%; Score 28; DB 2; Length 213;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsfxg 8
| | | |
Db 74 WKACAFAG 81

RESULT 8
A60032
Cerebellin-like glycoprotein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 31-Mar-2000
C:Accession: A60032
R:Wada, C.; Ohnani, H.
Brain Res. Mol. Brain Res. 9, 71-77, 1991
A:Title: Molecular cloning of rat cerebellin-like protein cDNA which encodes a novel mem
A:Accession: A60032
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-224 <WAD>
```

```
C:Superfamily: complement subcomponent C1q chain A; complement C1q carboxyl-terminal
C:Keywords: glycoprotein; transmembrane protein
F:32-49/Domain: transmembrane #status predicted <TM>
F:50-224/Domain: extracellular #status predicted <EXT>
F:94-223/Domain: complement C1q carboxyl-terminal homology <C10>
F:53,110/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match
Best Local Similarity 80.0%; Score 28; DB 2; Length 224;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsfxg 8
| | | |
Db 211 WKXSTFSG 218

RESULT 9
C96633
Probable Serine/Threonine protein kinase F8A5.29 [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C:Accession: C96633
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A66141; MUID:21016719
A:Accession: C96633
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-428 <STO>
A:Cross-references: GB:AE005173; NID:g2462748; PIDN:AB71967.1; GSPDB:GN00141
C:Genetics:
A:Gene: F8A5_29
A:Map position: 1
C:Superfamily: clathrin coat assembly protein AP50

Query Match
Best Local Similarity 80.0%; Score 28; DB 2; Length 428;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsfxg 8
| | | |
Db 346 WKIKSFG 353

RESULT 10
S65290
Clathrin-associated protein complex medium chain - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein P0394; protein YPL299c
C:Species: Saccharomyces cerevisiae
C:Date: 10-Dec-1994 #sequence_revision 31-May-1996 #text_change 20-Jun-2000
C:Accession: S65290; S19692; S17028
R:Messinguy, F.; Dubois, E.; Vlerendeels, F.; Scherens, B.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64935
A:Accession: S65290
A:Molecule type: DNA
A:Residues: 1-475 <MES>
A:Cross-references: EMBL:Z73615; NID:g1370535; PIDN:CAA97989.1; PID:g1370536; MIPS:YP
A:Experimental source: strain S288C (AB972)
R:Duesterhoft, A.; Floeth, M.; Frit, M.; Hilbert, H.; Moestl, D.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S65292
A:Accession: S65292
```

A:Molecule type: DNA
A:Residues: 305-475 <DUE>
A:Cross-references: EMBL:Z73615; MIPS:YPL259C
A:Experimental source: strain S288C (AB972)
R:Nakayama, Y.; Goebel, M.; O'Brine Greco, B.; Lemmon, S.; Pingchang Chow, E.; Kirchhauser, J. Biochem. 202, 563-574, 1991
A:Title: The medium chains of the mammalian clathrin-associated proteins have a homolog
A:Reference number: S19692; MUID:92104180
A:Accession: S19692
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-213, 'I', 215, 'H', 217-221, 'K', 223-421, 'S', 423-475 <NAK>
A:Cross-references: EMBL:X60288
R:Nakayama, Y.; Goebel, M.; O'Brine Greco, B.; Lemmon, S.; Pingchang Chow, E.; Kirchhauser, J. Biochem. 202, 563-574, 1991
A:Title: The medium chains of the mammalian clathrin associated proteins have an h
A:Reference number: S17028
A:Accession: S17028
A:Molecule type: DNA
A:Residues: 1-213, 'I', 215, 'H', 217-221, 'K', 223-432, 'R', 434-439, 'W', 441-449, 451-475 <NA2>
A:Cross-references: EMBL:X60288; NID:94799; PIDN:CAA42828.1; PID:94800
C:Genetics:
A:Gene: SGD:APM1; YAP54
A:Cross-references: SGD:S0006180; MIPS:YPL259C
A:Map position: 16L
C:Superfamily: clathrin coat assembly protein AP50
C:Keywords: transmembrane protein
F:65-81/Domain: transmembrane #status predicted <TMM>

Query Match 80.0%; Score 28; DB 2; Length 475;
Best Local Similarity 62.5%; Pred. No. 92;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxsfxg 8
|||
Db 382 WKIRSFPG 389

RESULT 11
A29947
prostaglandin-endoperoxide synthase (EC 1.14.99.1) precursor - sheep
N:Alternate names: prostaglandin endoperoxide synthetase
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 10-Dec-1999
C:Accession: A29947
R:Merlie, J.P.; Fagan, D.; Mudd, J.; Needleman, P.
J. Biol. Chem. 263, 3550-3553, 1988
A:Title: Isolation and characterization of the complementary DNA for sheep seminal vesic
A:Reference number: A29947; MUID:88153641
A:Accession: A29947
A:Molecule type: mRNA
A:Residues: 1-599 <MER>
A:Cross-references: GB:M18243; NID:q165843; PIDN:AAA31511.1; PID:q165844; GB:J03199
C:Superfamily: human prostaglandin-endoperoxide synthase; EGF homology
C:Keywords: oxidoreductase
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-599/Product: prostaglandin-endoperoxide synthase #status predicted <MAT>
F:35-68/Domain: EGF homology <EGF>

Query Match 80.0%; Score 28; DB 2; Length 599;
Best Local Similarity 50.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxsfxg 8
|||
Db 544 WKASTFCG 551

RESULT 12
A28960
prostaglandin-endoperoxide synthase (EC 1.14.99.1) G/H precursor - sheep

C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 10-Dec-1999
C:Accession: A28960
R:Dewilt, D.L.; Smith, W.L.
Proc. Natl. Acad. Sci. U.S.A. 85, 1412-1416, 1988
A:Title: Primary structure of prostaglandin G/H synthase from sheep vesicular gland d
A:Reference number: A28960; MUID:88144447
A:Accession: A28960
A:Molecule type: mRNA
A:Residues: 1-600 <DEW>
A:Cross-references: GB:J03599; NID:q166035; PIDN:AAA31576.1; PID:q166036
C:Superfamily: human prostaglandin-endoperoxide synthase; EGF homology
C:Keywords: oxidoreductase
F:36-69/Domain: EGF homology <EGF>

Query Match 80.0%; Score 28; DB 2; Length 600;
Best Local Similarity 50.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxsfxg 8
|||
Db 545 WKASTFCG 552

RESULT 13
S00561
prostaglandin-endoperoxide synthase (EC 1.14.99.1) precursor - sheep
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 10-Dec-1999
C:Accession: S00561
R:Yokoyama, C.; Takai, T.; Tanabe, T.
FEBS Lett. 231, 347-351, 1988
A:Title: Primary structure of sheep prostaglandin endoperoxide synthase deduced from
A:Reference number: S00561; MUID:88196421
A:Accession: S00561
A:Molecule type: mRNA
A:Residues: 1-600 <YOK>
A:Cross-references: EMBL:Y00750; NID:q1361; PIDN:CAA68719.1; PID:q1362
A:Note: part of this sequence, including the amino end of the mature protein, was con
A:Note: 97-His, 164-Gly, 436-Gln, 520-Gln, 520-Lys, and 525-Ile were also found
C:Superfamily: human prostaglandin-endoperoxide synthase; EGF homology
C:Keywords: oxidoreductase
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-600/Product: prostaglandin-endoperoxide synthase #status experimental <MAT>
F:36-63/Domain: EGF homology <EGF>

Query Match 80.0%; Score 28; DB 2; Length 600;
Best Local Similarity 50.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxsfxg 8
|||
Db 545 WKASTFCG 552

RESULT 14
S23164
light-harvesting protein alpha chain - Ectothiorhodospira halochloris
N:Alternate names: antenna pigment protein alpha chain
C:Species: Ectothiorhodospira halochloris
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 21-Aug-1998
C:Accession: S23164
R:Wagner-Huber, R.; Brunscholz, R.A.; Bissig, I.; Frank, G.; Suter, F.; Zuber, H.
Eur. J. Biochem. 205, 917-925, 1992
A:Title: The primary structure of the antenna polypeptides of Ectothiorhodospira halo
A:Reference number: S23164; MUID:92249336
A:Accession: S23164
A:Molecule type: protein
A:Residues: 1-65 <WAG>
C:Superfamily: light-harvesting protein alpha chain
C:Keywords: antenna complex; bacteriorhodophyll; blocked amino end; light-harvesting

A:Title: The primary structure of the antenna polypeptides of *Ecotrichorhodospira halochloris*
A:Reference number: S21164; MUID:92249336
A:Accession: S21164
A:Molecule type: protein
A:Residues: 1-65 <WAG>
C:Superfamily: Light-harvesting protein alpha chain
C:Keywords: antenna complex; bacteriorhodopsin; blocked amino end; light-harvesting protein; Modified site: N-formylmethionine #status experimental
F:1/Modified site: N-formylmethionine #status experimental

Query Match 77.8%; Score 28; DB 2; Length 65;
Best Local Similarity 50.0%; Pred. No. 11;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxyxg 8
| | | | |
Db 58 WRRSYDG 65

RESULT 12
S58121
Ig heavy chain V region precursor - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jan-2000
C:Accession: S58121
R:Kueppers, R.
Submitted to the EMBL Data Library, July 1995
A:Reference number: S58121
A:Accession: S58121
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-126 <KUE>
A:Cross-references: EMBL:X89595; NID:g929650; PIDN:CAA61756.1; PID:g929651
C:Genetics:
A:Introns: 9/1
C:Superfamily: Immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotrimer; immunoglobulin
F:27-110/Domain: Immunoglobulin homology <IMM>

Query Match 77.8%; Score 28; DB 2; Length 126;
Best Local Similarity 50.0%; Pred. No. 21;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxyxg 8
| | | | |
Db 114 WRRSYDG 121

RESULT 13
S21041
cytochrome-c oxidase (EC 1.9.3.1) chain II - *Sulfolobus acidocaldarius*
C:Species: *Sulfolobus acidocaldarius*
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 09-Sep-1997
C:Accession: S21041
R:Lucibben, M.; Kolmerer, B.; Saraste, M.
EMBO J. 11, 805-812, 1992
A:Title: An archaeobacterial terminal oxidase combines core structures of two mitochondria
A:Reference number: S21041; MUID:92192013
A:Accession: S21041
A:Molecule type: DNA
A:Residues: 1-168 <LUC>
A:Cross-references: EMBL:X62643; NID:g46681; PID:g46682
C:Genetics:
A:Gene: soxa
C:Keywords: copper binding; electron transfer; heme; membrane-associated complex; oxidoreductase

Query Match 77.8%; Score 28; DB 2; Length 168;
Best Local Similarity 50.0%; Pred. No. 28;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 wxxyxg 8

Db 147 WRDAEYAG 154

RESULT 14
H75258
Probable 3-demethylubiquinone-9-3-methyltransferase - *Deinococcus radiodurans* (strain C:Species: *Deinococcus radiodurans*
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: H75258
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Yamachyan, J.; Lam, P.; McDonald, L.; Uitterback, T.; Zalewski, C.; S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.
A:Reference number: A75250; MUID:20036896
A:Accession: H75258
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-206 <WHI>
A:Cross-references: GB:AE002085; GB:AE000513; NID:g6460383; PIDN:AAF12100.1; PID:g646
C:Experimental source: strain R1
C:Genetics:
A:Gene: DR2562
A:Map position: 1

Query Match 77.8%; Score 28; DB 2; Length 206;
Best Local Similarity 50.0%; Pred. No. 34;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxyxg 8
| | | | |
Db 111 WRRSYDG 118

RESULT 15
A35742
aqualysin (EC 3.4.21.-) I precursor - *Thermus aquaticus*
C:Species: *Thermus aquaticus*
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: A35742; S00620; S00324
R:Terada, I.; Kwon, S.T.; Miyata, Y.; Matsuzawa, H.; Ohta, T.
J. Biol. Chem. 265, 6576-6581, 1990
A:Title: Unique precursor structure of an extracellular protease, aqualysin I, with N
A:Reference number: A35742; MUID:90216674
A:Accession: A35742
A:Molecule type: DNA
A:Residues: 1-513 <TER>
A:Cross-references: GB:J90108; GB:D90108; GB:J05414; NID:g217171; PIDN:BAA14135.1; PI
R:Kwon, S.T.; Terada, I.; Matsuzawa, H.; Ohta, T.
Eur. J. Biochem. 173, 491-497, 1988
A:Title: Nucleotide sequence of the gene for aqualysin I (a thermophilic alkaline ser
A:Reference number: S00620; MUID:88225062
A:Accession: S00620
A:Molecule type: DNA
A:Residues: 75-442 <KWO>
A:Cross-references: EMBL:X07734; NID:g48069; PIDN:CAA30559.1; PID:g602091
A:Note: part of this sequence, including the amino and carboxyl ends of the mature pr
R:Matsuzawa, H.; Tokugawa, K.; Hamaoka, M.; Mizoguchi, M.; Taguchi, H.; Terada, I.; K
Eur. J. Biochem. 171, 441-447, 1988
A:Title: Purification and characterization of aqualysin I (a thermophilic alkaline se
A:Reference number: S00324; MUID:88151937
A:Accession: S00324
A:Molecule type: protein
A:Residues: 128-170 <MAT>
C:Superfamily: subtilisin; subtilisin homology
C:Keywords: extracellular protein; hydrolase; serine protease
F:1-14/Domain: signal sequence #status predicted <SIG>
F:115-127/Domain: propeptide #status predicted <PRO>
F:128-408/Product: aqualysin I #status experimental <MAT>
F:137-364/Domain: subtilisin homology <SBT>

F:58-74/Buain: transmembrane #status predicted <TM3>
F:85-101/Domain: transmembrane #status predicted <TM4>
F:127-143/Domain: transmembrane #status predicted <TM5>
F:182-198/Domain: transmembrane #status predicted <TM6>
F:213-229/Domain: transmembrane #status predicted <TM7>
F:240-256/Domain: transmembrane #status predicted <TM8>

Query Match 83.3%; Score 30; DB 2; Length 261;
Best Local Similarity 62.5%; Pred. No. 15;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxsyxg 8
|| || |
Db 28 WRRMSYFG 35

RESULT 7

E85798 hypothetical protein yebI [imported] - Escherichia coli (strain O157:H7)

C:Species: Escherichia coli

C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001

C:Accession: E85798

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: AB5480; MUID:21074935; PMID:11206551

A:Accession: E85798

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-261 <STO>

A:Cross-references: GB:AE005174; NID:g12515915; PIDN:AGS6849.1; GSPDB:GN00145; UMG:229

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: yebI

Query Match 83.3%; Score 30; DB 2; Length 261;
Best Local Similarity 62.5%; Pred. No. 15;

Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxsyxg 8
|| || |
Db 28 WRRMSYFG 35

RESULT 8

H70947 hypothetical protein RV3170 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: H70947

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellton, S.; Squares, S.
Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sultson, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987

A:Accession: H70947

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-448 <COL>

A:Cross-references: GB:AL021646; GB:AL123456; NID:g3242278; PIDN:CAI16635.1; PID:e124876

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: RV3170

Query Match 80.6%; Score 29; DB 2; Length 448;
Best Local Similarity 50.0%; Pred. No. 43;

Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 wrxxsyxg 8
|| || |
Db 306 WRASGYSG 313

RESULT 9

T04949 hypothetical protein F7J7.120 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999

C:Accession: T04949

R:Devan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.
submitted to the Protein Sequence Database, July 1998

A:Reference number: Z15391

A:Accession: T04949

A:Molecule type: DNA

A:Residues: 1-648 <BEV>

A:Cross-references: EMBL:AL021960

A:Experimental source: cultivar Columbia; BAC clone F7J7

C:Genetics:

A:Map position: 4

A:Introns: 44/3; 76/3; 128/3; 174/1; 289/3; 329/3; 358/3

A>Note: F7J7.120

Query Match 80.6%; Score 29; DB 2; Length 648;
Best Local Similarity 50.0%; Pred. No. 63;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 wrxxsyxg 8
|| || |
Db 201 WRSSKTYG 208

RESULT 10

F86812 phosphoketolase [imported] - Lactococcus lactis subsp. lactis (strain IL1403)

C:Species: Lactococcus lactis subsp. lactis

C>Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 23-Mar-2001

C:Accession: F86812

R:Bolecin, A.; Winkler, P.; Manger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; En
Genome Res. In press, 2001

A:Title: The complete genome sequence of the lactic acid bacterium.

A:Reference number: A86625

A:Accession: F86812

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-822 <STO>

A:Cross-references: GB:AE005176; NID:g12724499; PIDN:AAK05600.1; GSPDB:GN00146

A:Experimental source: strain IL1403

C:Genetics:

A:Gene: ptk

Query Match 80.6%; Score 29; DB 2; Length 822;
Best Local Similarity 50.0%; Pred. No. 80;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxsyxg 8
|| || |
Db 17 WRAATYIG 24

RESULT 11

S23164 light-harvesting protein alpha chain - Ectothiorhodospira halochloris

N:Alternate names: antenna pigment protein alpha chain

C:Species: Ectothiorhodospira halochloris

C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 21-Aug-1998

C:Accession: S23164

R:Wagner-Huber, R.; Brunisholz, R.A.; Bissig, I.; Frank, G.; Suter, F.; Zuber, H.
Eur. J. Biochem. 205, 917-925, 1992

hypothetical protein recB [Imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: G85933
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,
Mature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: G85933
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1180 <STO>
A:Cross-references: GB:AE005174; NID:g12517302; PIDN:AAG57931.1; GSPDB:GN00145; UWGP:Z41
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: recB
C:Superfamily: exodeoxyribonuclease V 135K chain

Query Match 86.1%; Score 31; DB 2; Length 1180;
Best Local Similarity 62.5%; Pred. No. 23;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 vrxxsyxg 8
|| || |
Db 901 WRVSYSG 908

RESULT 3
B82091
exodeoxyribonuclease V, 135 kDa chain VC2320 [Imported] - Vibrio cholerae (strain N16961)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: B82091
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Churchson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F.
L.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: B82091
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1208 <HEI>
A:Cross-references: GB:AE004303; GB:AE003852; NID:g9656890; PIDN:AAF95464.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC2320
A:Map position: 1
C:Superfamily: exodeoxyribonuclease V 135K chain

Query Match 88.9%; Score 32; DB 2; Length 1208;
Best Local Similarity 62.5%; Pred. No. 24;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 vrxxsyxg 8
|| || |
Db 906 WRVSYSG 913

RESULT 4
H71119
hypothetical protein PH0727 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 05-Nov-1999
C:Accession: H71119
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Okuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; MUID:98344137

A:Accession: H71119
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-278 <RAW>
A:Cross-references: GB:AP000003; NID:g3236130; PIDN:BAA29818.1; PID:d1030761; PID:g32
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBa
C:Genetics:
A:Gene: PH0727

Query Match 86.1%; Score 31; DB 2; Length 278;
Best Local Similarity 62.5%; Pred. No. 9.3;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 vrxxsyxg 8
|| || |
Db 99 WRVSYSG 106

RESULT 5
T16557
hypothetical protein K04E7.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T16557
R:Nhan, M.
submitted to the EMBL Data Library, October 1995
A:Description: The sequence of C. elegans cosmid K04E7.
A:Reference number: Z18535
A:Accession: T16557
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-551 <NHA>
A:Cross-references: EMBL:U39666; NID:g1049408; PID:g1049411; PIDN:AAA80412.1; CESP:KO
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:K04E7.3
A:Introns: 43/3; 81/3; 112/3; 153/1; 191/2; 213/3; 241/2; 281/3; 324/1; 364/1; 393/3;

Query Match 86.1%; Score 31; DB 2; Length 551;
Best Local Similarity 62.5%; Pred. No. 19;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 vrxxsyxg 8
|| || |
Db 376 WRVSYSG 383

RESULT 6
C64948
probable membrane protein yeb1 - Escherichia coli
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
C:Accession: C64948
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.: Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: C64948
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-261 <BLAT>
A:Cross-references: GB:AE000280; GB:U00096; NID:g1788163; PIDN:AAC74929.1; PID:g17881
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: yeb1
C:Superfamily: conserved hypothetical protein HI0360
C:Keywords: transmembrane protein; transport protein
F:11-27/Domain: transmembrane #status predicted <TM1>
F:38-54/Domain: transmembrane #status predicted <TM2>

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:37:28 ; Search time 63.57 Seconds
(without alignments)
9.586 Million cell updates/sec

Title: 09-185908-1e
Perfect score: 36
Sequence: 1 wrxsyxxg 8

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : PIR.68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	32	88.9	1180	1 NCECX5	exodeoxyribonuclease V (EC 3.1.11.5) 135K chain - Escherichia coli
2	32	88.9	1180	2 G85933	exodeoxyribonuclease V (EC 3.1.11.5) 135K chain - Escherichia coli
3	32	88.9	1208	2 B82091	exodeoxyribonuclease V (EC 3.1.11.5) 135K chain - Escherichia coli
4	31	86.1	278	2 H71119	hypothetical protein
5	31	86.1	551	2 T16557	hypothetical protein
6	30	83.3	261	2 C64948	probable membrane protein
7	30	83.3	261	2 C85798	hypothetical protein
8	29	80.6	448	2 H70947	hypothetical protein
9	29	80.6	648	2 T04949	hypothetical protein
10	29	80.6	822	2 F86812	phosphoketolase [lactate] from Escherichia coli
11	28	77.8	65	2 S23164	light-harvesting protein P680
12	28	77.8	126	2 S58121	cytochrome c oxidase subunit I from Escherichia coli
13	28	77.8	168	2 S21041	cytochrome c oxidase subunit II from Escherichia coli
14	28	77.8	206	2 H75258	probable 3-demethyladenosine methyltransferase from Escherichia coli
15	28	77.8	513	1 A35742	adenosine methyltransferase from Escherichia coli
16	28	77.8	669	1 S62919	probable membrane protein
17	28	77.8	700	2 T06088	hypothetical protein
18	28	77.8	988	2 H71338	conserved hypothetical protein
19	28	77.8	1289	2 T18212	parasporal crystal protein
20	27	75.0	116	1 Q4ADE2	early E4 11K protein
21	27	75.0	116	1 Q4ADE2	early E4 11K protein
22	27	75.0	154	2 E81795	probable tRNA-lysine synthetase from Escherichia coli
23	27	75.0	154	2 E81218	RNA methyltransferase from Escherichia coli
24	27	75.0	159	1 Q0CV2	aphid transmission factor from Escherichia coli
25	27	75.0	175	2 T11179	NADH dehydrogenase subunit 1 from Escherichia coli
26	27	75.0	187	2 H72503	hypothetical protein
27	27	75.0	199	2 F96737	hypothetical protein
28	27	75.0	241	2 B83447	hypothetical protein
29	27	75.0	246	2 T32510	hypothetical protein

30	27	75.0	260	2 F82120	zinc ABC transport protein
31	27	75.0	261	2 A64066	probable membrane protein
32	27	75.0	262	2 F82959	permease of ABC transporter
33	27	75.0	270	2 T26480	hypothetical protein
34	27	75.0	303	2 T00479	probable phosphatidyl transferase
35	27	75.0	323	2 D82987	hypothetical protein
36	27	75.0	325	2 T23426	hypothetical protein
37	27	75.0	325	2 T02325	probable DOP zinc
38	27	75.0	359	2 T02011	hypothetical protein
39	27	75.0	374	2 D72288	hypothetical protein
40	27	75.0	377	2 T01484	homeobox protein 1
41	27	75.0	383	2 T04641	hypothetical protein
42	27	75.0	441	2 G83212	hypothetical protein
43	27	75.0	455	2 E82658	beta-lactamase X1
44	27	75.0	458	2 T16123	hypothetical protein
45	27	75.0	474	1 C64801	ylea protein - Esc
46	27	75.0	474	1 F85566	hypothetical protein
47	27	75.0	531	2 T49058	hypothetical protein
48	27	75.0	636	2 H83040	biosynthetic argin
49	27	75.0	748	2 A56047	gamma-interferon a
50	27	75.0	851	2 A46160	interferon alpha-1

ALIGNMENTS

RESULT 1
NCECX5
exodeoxyribonuclease V (EC 3.1.11.5) 135K chain - Escherichia coli
N:Alternate names: exonuclease 135K polypeptide; recBC DNase 135K polypeptide
C:Species: Escherichia coli
C>Date: 31-Mar-1988 #sequence, revision 31-Mar-1988 #text, change 19-Jan-2001
C:Accession: A25532; E65064
R:Finch, P.W.; Storey, A.; Chapman, K.E.; Brown, K.; Hickson, I.D.; Emmerson, P.T.
Nucleic Acids Res. 14, 8573-8582, 1986
A:Title: Complete nucleotide sequence of the Escherichia coli recB gene.
A:Reference number: A25532; MUID:87066729
A:Accession: A25532
A:Molecule type: DNA
A:Residues: 1-1180 <FIN>
A:Cross-references: GB:X04581; NID:942680; PIDN:CA28250.1; PID:942682
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.: Rose, D.J.; Mau, B.; Shao, Y.
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: E65064
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1180 <BLAT>
A:Cross-references: GB:AE000365; GB:U00096; NID:92367163; PIDN:AACT5859.1; PID:917891
A:Experimental source: strain K-12, substrain MG1655
C:Comment: This enzyme is required for efficient DNA repair; it catalyzes the unwinding of these activities require concomitant hydrolysis of ATP.
C:Genetics:
A:Gene: recB
A:Map position: 61 min
C:Superfamily: exodeoxyribonuclease V 135K chain
C:Keywords: ATP; DNA repair; hydrolyase; nucleotide binding; P-loop
F:23-30/Region: nucleotide-binding motif A (P-loop)

Query Match 88.9%, Score 32; DB 1; Length 1180;
Best Local Similarity 62.5%; Pred. No. 23;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxsyxxg 8
Db 901 WRVTSYSG 908

RESULT 2
G85933

Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 WXXXXXG 8

DB 315 WKGARFEG 322

Search completed: January 14, 2002, 07:23:43
Job time: 74 sec

```
; INFORMATION FOR SEQ ID NO: 82:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 374
;   TYPE: amino acid
;   TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: protein
; ORIGINAL SOURCE:
; ORGANISM: human
; FEATURE: Alcohol dehydrogenase, Table 3 Column 3
; PUBLICATION INFORMATION:
;   AUTHORS:
;   AUTHORS: Joernvall, H.
;   AUTHORS: Persson, M.
;   AUTHORS: Jeffery, J.
;   TITLE: Alcohol dehydrogenases
;   JOURNAL: Proceedings of the National Academy of Sciences, USA
;   VOLUME: 78
;   PAGES: 4226-4230
;   DATE: 1981
; US-07-857-224B-82

Query Match          71.4%; Score 25; DB 2; Length 374;
Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 wkxxafxg 8
Db 314 WKGAIFGG 321

RESULT 19
US-07-857-224B-85
; Sequence 85, Application US/07857224B
; Patent No. 5958784
; GENERAL INFORMATION:
; APPLICANT: Benner, Steven A.
; TITLE OF INVENTION: Predicting Folded Structures of Proteins
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven A. Benner
; STREET: Hadlaubstrasse 151
; CITY: Zurich
; STATE: none
; COUNTRY: Switzerland
; ZIP: (note: this is an international post code) CH-8092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/857,224B
; FILING DATE: 03/25/92
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA: none
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (International) 41 1 632 2830
; TELEFAX: (International) 41 1 262 2437
; TELEX: none
; INFORMATION FOR SEQ ID NO: 85:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 374
;   TYPE: amino acid
;   TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: protein
; ORIGINAL SOURCE:
; ORGANISM: mouse
; FEATURE: Alcohol dehydrogenase, Table 3 Column 6
; PUBLICATION INFORMATION:
;   AUTHORS:
```

```
; AUTHORS: Joernvall, H.
; AUTHORS: Persson, M.
; AUTHORS: Jeffery, J.
; TITLE: Alcohol dehydrogenases
; JOURNAL: Proceedings of the National Academy of Sciences, USA
; VOLUME: 78
; PAGES: 4226-4230
; DATE: 1981
; US-07-857-224B-85

Query Match          71.4%; Score 25; DB 2; Length 374;
Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 wkxxafxg 8
Db 314 WKGAIFGG 321

RESULT 20
US-07-857-224B-86
; Sequence 86, Application US/07857224B
; Patent No. 5958784
; GENERAL INFORMATION:
; APPLICANT: Benner, Steven A.
; TITLE OF INVENTION: Predicting Folded Structures of Proteins
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven A. Benner
; STREET: Hadlaubstrasse 151
; CITY: Zurich
; STATE: none
; COUNTRY: Switzerland
; ZIP: (note: this is an international post code) CH-8092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/857,224B
; FILING DATE: 03/25/92
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA: none
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (International) 41 1 632 2830
; TELEFAX: (International) 41 1 262 2437
; TELEX: none
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 375
;   TYPE: amino acid
;   TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: protein
; ORIGINAL SOURCE:
; ORGANISM: rat
; FEATURE: Alcohol dehydrogenase, Table 3 Column 7
; PUBLICATION INFORMATION:
;   AUTHORS:
;   AUTHORS: Joernvall, H.
;   AUTHORS: Persson, M.
;   AUTHORS: Jeffery, J.
;   TITLE: Alcohol dehydrogenases
;   JOURNAL: Proceedings of the National Academy of Sciences, USA
;   VOLUME: 78
;   PAGES: 4226-4230
;   DATE: 1981
; US-07-857-224B-86

Query Match          71.4%; Score 25; DB 2; Length 375;
```

```

: APPLICANT: Benner, Steven A.
: TITLE OF INVENTION: Predicting Folded Structures of Proteins
: NUMBER OF SEQUENCES: 114
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Steven A. Benner
: STREET: Hadlaubstrasse 151
: CITY: Zurich
: STATE: none
: COUNTRY: Switzerland
: ZIP: (note: this is an international post code) CH-8092
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
: OPERATING SYSTEM: Macintosh 7.0
: SOFTWARE: Microsoft Word
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/857,224B
: FILING DATE: 03/25/92
: CLASSIFICATION: 436
: PRIOR APPLICATION DATA: none
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (International) 41 1 632 2830
: TELEFAX: (International) 41 1 262 2437
: TELEX: none
: INFORMATION FOR SEQ ID NO: 80:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 374
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: ORIGINAL SOURCE:
: ORGANISM: horse
: FEATURE: Alcohol dehydrogenase, Table 3 Column 1
: PUBLICATION INFORMATION:
: AUTHORS: Joernvall, H.
: AUTHORS: Persson, M.
: AUTHORS: Jeffery, J.
: TITLE: Alcohol dehydrogenases
: JOURNAL: Proceedings of the National Academy of Sciences, USA
: VOLUME: 78
: PAGES: 4226-4230
: DATE: 1981
: US-07-857-224B-80

Query Match      71.4%; Score 25; DB 2; Length 374;
Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 wkxxafgx 8
Db 314 WKGAIFGC 321

RESULT 17
US-07-857-224B-81
: Sequence 81, Application US/07857224B
: Patent No. 5958784
: GENERAL INFORMATION:
: APPLICANT: Benner, Steven A.
: TITLE OF INVENTION: Predicting Folded Structures of Proteins
: NUMBER OF SEQUENCES: 114
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Steven A. Benner
: STREET: Hadlaubstrasse 151
: CITY: Zurich
: STATE: none
: COUNTRY: Switzerland
: ZIP: (note: this is an international post code) CH-8092
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
```

```

: COMPUTER: Apple Macintosh
: OPERATING SYSTEM: Macintosh 7.0
: SOFTWARE: Microsoft Word
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/857,224B
: FILING DATE: 03/25/92
: CLASSIFICATION: 436
: PRIOR APPLICATION DATA: none
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (International) 41 1 632 2830
: TELEFAX: (International) 41 1 262 2437
: TELEX: none
: INFORMATION FOR SEQ ID NO: 81:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 374
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: DESCRIPTION: protein
: ORIGINAL SOURCE:
: ORGANISM: horse
: FEATURE: Alcohol dehydrogenase, Table 3 Column 2
: PUBLICATION INFORMATION:
: AUTHORS: Joernvall, H.
: AUTHORS: Persson, M.
: AUTHORS: Jeffery, J.
: TITLE: Alcohol dehydrogenases
: JOURNAL: Proceedings of the National Academy of Sciences, USA
: VOLUME: 78
: PAGES: 4226-4230
: DATE: 1981
: US-07-857-224B-81

Query Match      71.4%; Score 25; DB 2; Length 374;
Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 wkxxafgx 8
Db 314 WKGAIFGC 321

RESULT 18
US-07-857-224B-82
: Sequence 82, Application US/07857224B
: Patent No. 5958784
: GENERAL INFORMATION:
: APPLICANT: Benner, Steven A.
: TITLE OF INVENTION: Predicting Folded Structures of Proteins
: NUMBER OF SEQUENCES: 114
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Steven A. Benner
: STREET: Hadlaubstrasse 151
: CITY: Zurich
: STATE: none
: COUNTRY: Switzerland
: ZIP: (note: this is an international post code) CH-8092
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
: OPERATING SYSTEM: Macintosh 7.0
: SOFTWARE: Microsoft Word
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/857,224B
: FILING DATE: 03/25/92
: CLASSIFICATION: 436
: PRIOR APPLICATION DATA: none
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (International) 41 1 632 2830
: TELEFAX: (International) 41 1 262 2437
: TELEX: none
```


DB 209 WKXSTFSG 216

RESULT 8

US-09-222-817-2

Sequence 2, Application US/09222817

Patent No. 6037154

GENERAL INFORMATION:

APPLICANT: MIKIKO SUGA, Masakazu SUGIMOTO, Tsuyoshi OSUMI, Tsuyoshi NAKAMATSU,

TITLE OF INVENTION: METHOD OF PRODUCING L-SERINE BY FERMENTATION

FILE REFERENCE: OP813

CURRENT APPLICATION NUMBER: US/09/222,817

CURRENT FILING DATE: 1998-12-30

EARLIER APPLICATION NUMBER: JP 10-3751

EARLIER FILING DATE: 1998-01-12

EARLIER APPLICATION NUMBER: JP 10-353521

EARLIER FILING DATE: 1998-12-11

NUMBER OF SEQ ID NOS: 18

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 2

LENGTH: 345

TYPE: PRF

ORGANISM: Corynebacterium glutamicum

US-09-222-817-2

Query Match 77.1%; Score 27; DB 3; Length 345;

Best Local Similarity 50.0%; Pred. No. 1.2e+02;

Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8

DB 132 WKRSSFNG 139

RESULT 9

US-09-222-786-2

Sequence 2, Application US/09222786A

Patent No. 6258573

GENERAL INFORMATION:

APPLICANT: MIKIKO SUGA, Masakazu SUGIMOTO, Tsuyoshi OSUMI, Tsuyoshi NAKAMATSU,

TITLE OF INVENTION: METHOD OF PRODUCING L-SERINE BY FERMENTATION

FILE REFERENCE: OP812

CURRENT APPLICATION NUMBER: US/09/222,786A

CURRENT FILING DATE: 1998-12-30

EARLIER APPLICATION NUMBER: JP 10-3751

EARLIER FILING DATE: 1998-01-12

EARLIER APPLICATION NUMBER: JP 10-353513

EARLIER FILING DATE: 1998-12-11

NUMBER OF SEQ ID NOS: 14

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 2

LENGTH: 345

TYPE: PRF

ORGANISM: Corynebacterium glutamicum

US-09-222-786-2

Query Match 77.1%; Score 27; DB 4; Length 345;

Best Local Similarity 50.0%; Pred. No. 1.2e+02;

Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8

DB 132 WKRSSFNG 139

RESULT 10

US-09-222-817-12

Sequence 12, Application US/09222817

Patent No. 6037154

GENERAL INFORMATION:

APPLICANT: MIKIKO SUGA, Masakazu SUGIMOTO, Tsuyoshi OSUMI, Tsuyoshi NAKAMATSU,

TITLE OF INVENTION: METHOD OF PRODUCING L-SERINE BY FERMENTATION

FILE REFERENCE: OP813

CURRENT APPLICATION NUMBER: US/09/222,817

CURRENT FILING DATE: 1998-12-30

EARLIER APPLICATION NUMBER: JP 10-3751

EARLIER FILING DATE: 1998-01-12

EARLIER APPLICATION NUMBER: JP 10-353521

EARLIER FILING DATE: 1998-12-11

NUMBER OF SEQ ID NOS: 18

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 12

LENGTH: 530

TYPE: PRF

ORGANISM: Brevibacterium flavum

US-09-222-817-12

Query Match 77.1%; Score 27; DB 3; Length 530;

Best Local Similarity 50.0%; Pred. No. 1.8e+02;

Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8

DB 132 WKRSSFNG 139

RESULT 11

US-09-222-817-14

Sequence 14, Application US/09222817

Patent No. 6037154

GENERAL INFORMATION:

APPLICANT: MIKIKO SUGA, Masakazu SUGIMOTO, Tsuyoshi OSUMI, Tsuyoshi NAKAMATSU,

TITLE OF INVENTION: METHOD OF PRODUCING L-SERINE BY FERMENTATION

FILE REFERENCE: OP813

CURRENT APPLICATION NUMBER: US/09/222,817

CURRENT FILING DATE: 1998-12-30

EARLIER APPLICATION NUMBER: JP 10-3751

EARLIER FILING DATE: 1998-01-12

EARLIER APPLICATION NUMBER: JP 10-353521

EARLIER FILING DATE: 1998-12-11

NUMBER OF SEQ ID NOS: 18

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 14

LENGTH: 530

TYPE: PRF

ORGANISM: Brevibacterium flavum

US-09-222-817-14

Query Match 77.1%; Score 27; DB 3; Length 530;

Best Local Similarity 50.0%; Pred. No. 1.8e+02;

Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8

DB 132 WKRSSFNG 139

RESULT 12

US-09-222-786-12

Sequence 12, Application US/09222786A

Patent No. 6258573

GENERAL INFORMATION:

APPLICANT: MIKIKO SUGA, Masakazu SUGIMOTO, Tsuyoshi OSUMI, Tsuyoshi NAKAMATSU,

TITLE OF INVENTION: METHOD OF PRODUCING L-SERINE BY FERMENTATION

FILE REFERENCE: OP812

CURRENT APPLICATION NUMBER: US/09/222,786A

CURRENT FILING DATE: 1998-12-30

EARLIER APPLICATION NUMBER: JP 10-3751

EARLIER FILING DATE: 1998-01-12

EARLIER APPLICATION NUMBER: JP 10-353513

EARLIER FILING DATE: 1998-12-11

NUMBER OF SEQ ID NOS: 14

OY 1 wxxxxfxg 8
Db 316 WKSAFAG 323

RESULT 5

US-08-637-759B-89
Sequence 89, Application US/08637759B
Patent No. 5876931
GENERAL INFORMATION:
APPLICANT: David William Holden
TITLE OF INVENTION: Identification of Genes
NUMBER OF SEQUENCES: 501
CORRESPONDENCE ADDRESS: 501
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/637,759B
FILING DATE: 03-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/02875
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPKS 101
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8794
INFORMATION FOR SEQ ID NO: 89:
SEQUENCE CHARACTERISTICS:
LENGTH: 759 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-637-759B-89

Query Match 82.9%; Score 29; DB 2; Length 759;
Best Local Similarity 62.5%; Pred. No. 96;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxfxg 8
Db 30 WKPAFAG 37

RESULT 6
US-08-871-355A-89
Sequence 89, Application US/08871355A
Patent No. 6015669
GENERAL INFORMATION:
APPLICANT: David William Holden
TITLE OF INVENTION: Identification of Genes
NUMBER OF SEQUENCES: 501
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center

STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/871,355A
FILING DATE: 09-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/02875
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPKS 101 CON
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8794
INFORMATION FOR SEQ ID NO: 89:
SEQUENCE CHARACTERISTICS:
LENGTH: 759 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-871-355A-89

Query Match 82.9%; Score 29; DB 3; Length 759;
Best Local Similarity 62.5%; Pred. No. 96;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxfxg 8
Db 30 WKPAFAG 37

RESULT 7
US-09-140-804-7
Sequence 7, Application US/09140804
Patent No. 6197930
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
FILE REFERENCE: 97-49
CURRENT APPLICATION NUMBER: US/09/140,804
CURRENT FILING DATE: 1998-08-26
EARLIER APPLICATION NUMBER: 60/056,983
EARLIER FILING DATE: 1997-08-26
NUMBER OF SEQ ID NOS: 47
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 7
LENGTH: 222
TYPE: PRT
ORGANISM: Rattus norvegicus
US-09-140-804-7

Query Match 77.1%; Score 27; DB 4; Length 222;
Best Local Similarity 50.0%; Pred. No. 81;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wxxxxfxg 8
Db 30 WKPAFAG 37

Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxfxg 8
|| ||
Db 316 WKSAFGG 323

RESULT 2
US-08-922-182-2
; Sequence 2, Application US/08922182
; Patent No. 5834300
; GENERAL INFORMATION:
; APPLICANT: Donohue, Timothy J
; APPLICANT: Barber, Robert D
; APPLICANT: Wiltuhn, Vernon
; TITLE OF INVENTION: MICROBIAL SYSTEM FOR FORMALDEHYDE
; TITLE OF INVENTION: SENSING AND REMEDIATION
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/922,182
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/608,241
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 960296.93511
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 376 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-922-182-2

Query Match 85.7%; Score 30; DB 2; Length 376;
Best Local Similarity 62.5%; Pred. No. 31;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxfxg 8
|| ||
Db 316 WKSAFGG 323

RESULT 3
US-08-919-953-2
; Sequence 2, Application US/08919953
; Patent No. 5837481
; GENERAL INFORMATION:
; APPLICANT: Donohue, Timothy J
; APPLICANT: Barber, Robert D
; APPLICANT: Wiltuhn, Vernon
; TITLE OF INVENTION: MICROBIAL SYSTEM FOR FORMALDEHYDE
; TITLE OF INVENTION: SENSING AND REMEDIATION
; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/919,953
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/608,241
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 960296.93511
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 376 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-919-953-2

Query Match 85.7%; Score 30; DB 2; Length 376;
Best Local Similarity 62.5%; Pred. No. 31;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxfxg 8
|| ||
Db 316 WKSAFGG 323

RESULT 4
US-09-192-983-2
; Sequence 2, Application US/09192983A
; Patent No. 6242244
; GENERAL INFORMATION:
; APPLICANT: Donohue, Timothy
; APPLICANT: Barber, Robert
; APPLICANT: Wiltuhn, Vernon
; TITLE OF INVENTION: Microbial System for Formaldehyde Sensing and
; TITLE OF INVENTION: Remediation
; FILE REFERENCE: 960296.95505
; CURRENT APPLICATION NUMBER: US/09/192,983A
; EARLIER FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: 08/919,953
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 08/608,241
; EARLIER FILING DATE: 1996-02-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 2
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Rhodobacter sphaeroides
US-09-192-983-2

Query Match 85.7%; Score 30; DB 4; Length 376;
Best Local Similarity 62.5%; Pred. No. 31;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:23:43 ; Search time 48.05 Seconds
(without alignments)
3.747 Million cell updates/sec

Title: 09-185908-1D
Perfect score: 35
Sequence: 1 wxxxxafxyg 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : Issued Patents.AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCRTUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	85.7	376	1 US-08-608-241-2	Sequence 2, Appli
2	30	85.7	376	2 US-08-922-182-2	Sequence 2, Appli
3	30	85.7	376	2 US-08-919-953-2	Sequence 2, Appli
4	30	85.7	376	4 US-09-192-983-2	Sequence 2, Appli
5	29	82.9	759	2 US-08-637-759B-89	Sequence 89, Appl
6	29	82.9	759	3 US-08-871-355A-89	Sequence 89, Appl
7	27	77.1	222	4 US-09-140-804-7	Sequence 7, Appli
8	27	77.1	345	3 US-09-222-817-2	Sequence 2, Appli
9	27	77.1	345	4 US-09-222-86-2	Sequence 2, Appli
10	27	77.1	530	3 US-09-222-817-12	Sequence 12, Appl
11	27	77.1	530	3 US-09-222-817-14	Sequence 14, Appl
12	27	77.1	530	4 US-09-222-786-12	Sequence 12, Appl
13	27	77.1	530	4 US-09-222-786-14	Sequence 14, Appl
14	25	71.4	214	4 US-09-214-278-1	Sequence 1, Appli
15	25	71.4	277	4 US-08-567-375-16	Sequence 16, Appl
16	25	71.4	374	2 US-07-857-224B-80	Sequence 80, Appl
17	25	71.4	374	2 US-07-857-224B-81	Sequence 81, Appl
18	25	71.4	374	2 US-07-857-224B-82	Sequence 82, Appl
19	25	71.4	374	2 US-07-857-224B-85	Sequence 85, Appl
20	25	71.4	375	2 US-07-857-224B-86	Sequence 86, Appl
21	25	71.4	544	2 US-08-587-680A-25	Sequence 25, Appl
22	25	71.4	592	3 US-08-991-813-2	Sequence 2, Appli
23	25	71.4	604	1 US-08-064-271-10	Sequence 10, Appl
24	25	71.4	604	1 US-08-487-753-2	Sequence 2, Appli
25	25	71.4	604	1 US-08-487-753-4	Sequence 4, Appli
26	25	71.4	604	1 US-08-487-753-5	Sequence 5, Appli
27	25	71.4	604	2 US-08-480-065-2	Sequence 2, Appli

28	25	71.4	604	2 US-08-480-065-4	Sequence 4, Appli
29	25	71.4	604	2 US-08-480-065-5	Sequence 5, Appli
30	25	71.4	604	3 US-08-487-744-2	Sequence 2, Appli
31	25	71.4	604	3 US-08-487-744-5	Sequence 4, Appli
32	25	71.4	604	3 US-08-487-744-5	Sequence 4, Appli
33	25	71.4	604	3 US-08-930-589A-18	Sequence 18, Appl
34	25	71.4	604	5 PCT-US93-09167-2	Sequence 2, Appli
35	25	71.4	604	5 PCT-US93-09167-4	Sequence 4, Appli
36	25	71.4	604	5 PCT-US93-09167-5	Sequence 5, Appli
37	25	71.4	980	2 US-08-473-553A-6	Sequence 6, Appli
38	25	71.4	985	2 US-08-473-553A-2	Sequence 2, Appli
39	25	71.4	1055	4 US-09-214-278-2	Sequence 2, Appli
40	25	71.4	1065	4 US-08-400-159-8	Sequence 8, Appli
41	25	71.4	1079	3 US-09-058-489-22	Sequence 22, Appl
42	25	71.4	1148	4 US-08-882-046-4	Sequence 4, Appli
43	25	71.4	1212	4 US-09-214-278-3	Sequence 3, Appli
44	25	71.4	1238	4 US-09-214-278-5	Sequence 5, Appli
45	25	71.4	1240	3 US-09-058-489-23	Sequence 23, Appl
46	25	71.4	1240	3 US-08-882-046-6	Sequence 6, Appli
47	25	71.4	1257	4 US-08-611-729A-8	Sequence 8, Appli
48	25	71.4	1347	3 US-09-058-489-24	Sequence 24, Appl
49	25	71.4	2987	2 US-08-970-269A-29	Sequence 29, Appl
50	25	71.4	2987	4 US-09-407-562-29	Sequence 29, Appl

ALIGNMENTS

RESULT 1

US-08-608-241-2
; Sequence 2, Application US/08608241

; Patent No. 5747328

; GENERAL INFORMATION:

; APPLICANT: Donohue, Timothy J

; APPLICANT: Barber, Robert D

; APPLICANT: Withuhn, Vernon

; TITLE OF INVENTION: MICROBIAL SYSTEM FOR FORMALDEHYDE

; TITLE OF INVENTION: SENSING AND REMEDIATION

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Quarles & Brady

; STREET: 1 South Pinckney Street

; CITY: Madison

; STATE: WI

; COUNTRY: US

; ZIP: 53703

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/608,241

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Seay, Nicholas J

; REGISTRATION NUMBER: 27,386

; REFERENCE/DOCKET NUMBER: 960296.93511

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 608-251-5000

; TELEFAX: 608-251-9166

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 376 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-608-241-2

Query Match 85.7%; Score 30; DB 1; Length 376;
Best Local Similarity 62.5%; Pred. No. 31;

PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149729.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.

PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 85.7%; Score 30; DB 21; Length 255;
Best Local Similarity 62.5%; Pred. No. 77;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxxxafxg 8
DB 193 wkqtafgg 200

Search completed: January 14, 2002, 07:36:00
Job time: 811 sec

PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 85.7%; Score 30; DB 21; Length 254;
Best Local Similarity 62.5%; Pred. No. 76;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxfxg 8
DB 192 wkyxatfg 199

RESULT 20

ACGI6747
ID AAGI6747 standard; Protein; 255 AA.

XX AAGI6747;

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 17510.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hydridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.

XX Arabidopsis thaliana.

OS
PN EPI033405-A2.

XX
PD 06-SEP-2000.

XX
PE 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130049.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-01332407.
PR 05-MAY-1999; 99US-01332484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140605.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145006.
PR 21-JUL-1999; 99US-0145008.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145146.
PR 23-JUL-1999; 99US-0145218.

PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0156589.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.

DR N-PSDB: AAC91570.
XX
XX Twenty eight nucleic acids encoding PRO polypeptides which are useful
PT for treating various tumors, e.g. breast cancer, and other
PT inflammatory, angiogenic and immunological disorders -
XX
XX Claim 31: Fig 36; 188pp; English.
XX
CC The present sequence is one of twenty eight novel PRO polypeptides. The
CC PRO polypeptides and their agonists, including antibodies, peptides, and
CC small molecule agonists, may be used to treat various tumors, e.g.,
CC cancers such as breast cancer, ovarian cancer, renal cancer, colorectal
CC cancer, uterine cancer, prostate cancer, lung cancer, bladder cancer,
CC central nervous system cancer, melanoma or leukaemia. They are also
CC useful for treating other disorders such as neuronal, glial, astrocytal,
CC hypothalamic and other glandular, macrophagal, epithelial, stromal and
CC blastocelial disorders, and inflammatory, angiogenic and immunological
CC disorders.
XX
XX Sequence 220 AA;
SQ

Query Match 85.7%; Score 30; DB 22; Length 220;
Best Local Similarity 62.5%; Pred. No. 67;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxafxg 8
Db 30 wkvtafifg 37

RESULT 18
AAM41723 standard; Protein: 235 AA.
ID AAM41723;
AC AAM41723;
XX
XX 22-OCT-2001 (first entry)
DT
XX
XX Human polypeptide SEQ ID NO 6654.
DE
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
XX Homo sapiens.
OS
XX
XX WO200153312-A1.
PN
XX
XX 26-JUL-2001.
PD
XX
XX 26-DEC-2000; 2000WO-US34263.
PF
XX
XX 21-JAN-2000; 2000US-0488725.
PR
XX 25-APR-2000; 2000US-0552317.
PR
XX 09-JUL-2000; 2000US-0598042.
PR
XX 19-JUL-2000; 2000US-0620312.
PR
XX 03-AUG-2000; 2000US-0653450.
PR
XX 14-SEP-2000; 2000US-0662191.
PR
XX 19-OCT-2000; 2000US-0692036.
PR
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Dirmnac RT;
XX
XX WPI: 2001-442253/47.
DR N-PSDB: AAI60879.

XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 2; SEQ ID NO 6654; 10078pp; English.
PS
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 235 AA;
SQ

Query Match 85.7%; Score 30; DB 22; Length 235;
Best Local Similarity 62.5%; Pred. No. 71;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxafxg 8
Db 45 wkvtafifg 52

RESULT 19
AAG17084 standard; Protein: 254 AA.
ID AAG17084;
AC AAG17084;
XX
XX 17-OCT-2000 (first entry)
DT
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 17971.
DE
XX
XX Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
KW
XX
XX Arabidopsis thaliana.
OS
XX
XX EP1033405-A2.
PN
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825.
PR
XX 05-MAR-1999; 99US-0123180.
PR
XX 09-MAR-1999; 99US-0123548.
PR
XX 23-MAR-1999; 99US-0125788.
PR
XX 25-MAR-1999; 99US-0126264.
PR
XX 29-MAR-1999; 99US-0126785.
PR
XX 01-APR-1999; 99US-0127462.
PR
XX 06-APR-1999; 99US-0128234.
PR
XX 08-APR-1999; 99US-0128714.
PR
XX 16-APR-1999; 99US-0129845.
PR
XX 19-APR-1999; 99US-0130077.
PR
XX 21-APR-1999; 99US-0130449.
PR
XX 23-APR-1999; 99US-0130510.
PR
XX 28-APR-1999; 99US-0130891.
PR
XX 30-APR-1999; 99US-0131449.
PR
XX 99US-0132048.

proteins they encode may be used in the prevention, treatment and diagnosis of diseases associated with inappropriate secretory protein/membrane protein expression. The nucleic acids and complementary sequences may also be used as DNA probes in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect and quantitate the presence of similar nucleic acid sequences in samples. They may also be used to study the expression and function of secretory proteins/membrane polypeptides and their role in metabolism. The polypeptides may be used as antigens in the production of antibodies against them and in assays to identify modulators (agonists and antagonists) of expression and activity. The antibodies and antagonists may also be used as therapeutic agents to down regulate expression and activity. The antibodies may also be used as diagnostic agents for detecting the presence of the polypeptides in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). Examples of diseases which may be treated include rheumatoid arthritis and diabetes.

Sequence 220 AA;

Query Match 85.7%; Score 30; DB 22; Length 220;
Best Local Similarity 62.5%; Pred. No. 67;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 wkxxafxg 8
|| || |
30 wkvtatfg 37

Db

RESULT 16
AAB6183
ID AAB6183 standard; protein; 220 AA.
XX
AC AAB6183;
XX
DT 02-APR-2001 (first entry)
XX
DE Protein of the invention #95.
XX
KW Secreted; transmembrane; gene therapy.
XX
OS Unidentified.
XX
PN WO200078961-A1.
XX
PD 28-DEC-2000.
XX
PF 18-FEB-2000; 2000WO-US04342.
XX
PR 23-JUN-1999; 99US-0141037.
PR 20-JUL-1999; 99US-0144758.
PR 26-JUL-1999; 99US-0145698.
PR 01-SEP-1999; 99WO-US20111.
PR 29-OCT-1999; 99US-0162506.
PR 30-NOV-1999; 99WO-US28313.
PR 02-DEC-1999; 99WO-US28551.
PR 16-DEC-1999; 99WO-US30095.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00376.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;
PI Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
PI Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;
PI Watanabe CK, Williams PM, Wood WI;
XX
DR WPI; 2001-071395/08.
XX
PT Secreted and transmembrane proteins and nucleic acids designated PRO,
PT useful as hybridization probes, in chromosome and gene mapping and gene
XX therapy -
XX

Claim 1; Fig 190; 787pp; English.

The present invention relates to secreted and transmembrane proteins. These proteins and the DNA encoding them may be used as hybridization probes, in chromosome and gene mapping and in the generation of anti-sense RNA and DNA. They may also be used to generate either transgenic animals or knockout animals which are in turn useful for development and screening of therapeutically useful reagents. The nucleic acids may also be used in gene therapy.

Sequence 220 AA;

Query Match 85.7%; Score 30; DB 22; Length 220;
Best Local Similarity 62.5%; Pred. No. 67;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 wkxxafxg 8
|| || |
30 wkvtatfg 37

Db

RESULT 17
AAB50968
ID AAB50968 standard; Protein; 220 AA.
XX
AC AAB50968;
XX
DT 21-MAR-2001 (first entry)
XX
DE Human PRO1488 protein.
XX
KW Human; PRO; cytosolic; neotropic; neuroprotective; respiratory general;
KW antiinflammatory; antiangiogenic; immunosuppressive; immunostimulant;
KW PRO agonist; cancer; inflammatory disorder; immunological disorder.
XX
OS Homo sapiens.
XX
PN WO200073348-A2.
XX
PD 07-DEC-2000.
XX
PF 30-MAY-2000; 2000WO-US14941.
XX
PR 02-JUN-1999; 99WO-US12252.
PR 22-JUN-1999; 99US-0140650.
PR 23-JUN-1999; 99US-0141037.
PR 20-JUL-1999; 99US-0144758.
PR 01-SEP-1999; 99WO-US20111.
PR 08-SEP-1999; 99WO-US20594.
PR 29-OCT-1999; 99US-0162506.
PR 30-NOV-1999; 99WO-US28313.
PR 01-DEC-1999; 99WO-US28634.
PR 02-DEC-1999; 99WO-US28551.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30999.
PR 06-JAN-2000; 2000WO-US00376.
PR 11-FEB-2000; 2000WO-US03565.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 02-MAR-2000; 2000WO-US05841.
PR 10-MAR-2000; 2000US-0187202.
PR 15-MAR-2000; 2000WO-US06319.
PR 30-MAR-2000; 2000WO-US06884.
PR 17-MAY-2000; 2000WO-US13705.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker KP, Goddard A, Gurney AL, Hebert C, Henzel W, Kabakoff RC;
PI Shelton DL, Smith V, Watanabe CK, Wood WI;
XX
DR WPI; 2001-016509/02.
XX

PS Claim 12; Fig 190; 773pp; English.

CC AAA37022 to AAA37144 encode the new isolated human transmembrane,
CC receptor or secreted PRO polypeptides given in AAy99340 to AAy9462. The
CC transmembrane and receptor PRO proteins can be used for screening of
CC potential peptide or small molecule inhibitors of the relevant
CC receptor/ligand interactions. The polypeptides and nucleotide sequences
CC encoding then have various industrial applications, including uses as
CC pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
CC PCR primers and hybridisation probes used in the isolation of the PRO
CC polypeptides from the present invention.

CC Sequence 220 AA:

Query Match 85.7%; Score 30; DB 21; Length 220;
Best Local Similarity 62.5%; Pred. No. 67;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8
|| || |
DB 30 wkvtatfig 37

RESULT 14
ID AAA39937 standard; Protein; 220 AA.

AC AAA39937;
XX
XX 22-OCT-2001 (first entry)
XX
XX Human polypeptide SEQ ID NO 3082.
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukemia.

OS Homo sapiens.
XX
XX WO200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.

XX (HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
XX N-PSDB; AAI59093.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 4; SEQ ID NO 3082; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and

CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

CC Sequence 220 AA:

Query Match 85.7%; Score 30; DB 22; Length 220;
Best Local Similarity 62.5%; Pred. No. 67;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8
|| || |
DB 30 wkvtatfig 37

RESULT 15
ID AAB88319 standard; Protein; 220 AA.

AC AAB88319;
XX
XX 23-MAY-2001 (first entry)
XX
XX Human membrane or secretory protein clone PSEC0005.
XX
XX Human; secretory protein; membrane protein; vaccine; gene therapy;
KW rheumatoid arthritis; diabetes.

OS Homo sapiens.
XX
XX EP1067182-A2.
XX
XX 10-JAN-2001.
XX
XX 07-JUL-2000; 2000EP-0114090.
XX
XX 08-JUL-1999; 99JP-0194179.
XX 11-JAN-2000; 2000JP-0118775.
XX 02-MAY-2000; 2000JP-0183766.

XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
PI WPI: 2001-093989/11.
XX N-PSDB; AAF93746.
XX
XX Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT gene therapy or as candidate target molecules in drug development -
XX
XX Claim 1; SEQ ID 6; 609pp + CD ROM; English.

CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC which encode human secretory or membrane proteins represented by
CC AAB88317 - AAB88419. Included in the invention are primers
CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC cDNA sequences of the invention. The invention also includes methods for
CC the production of antibodies directed against the proteins, and cDNA
CC sequences, which can be used in vaccines. The polynucleotide sequences
CC can be used in gene therapy. The polynucleotide sequences and the

```
XX      .
KW      Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
XX      Transmembrane; secretion; immunoadhesion; pharmaceutical; screening.
OS      Homo sapiens.
XX      WO200012708-A2.
XX      09-MAR-2000.
XX      01-SEP-1999;      99WO-US20111.
XX      01-SEP-1998;      98US-0098716.
PR      01-SEP-1998;      98US-0098749.
PR      01-SEP-1998;      98US-0098750.
PR      02-SEP-1998;      98US-0098803.
PR      02-SEP-1998;      98US-0098821.
PR      02-SEP-1998;      98US-0098843.
PR      09-SEP-1998;      98US-0099536.
PR      09-SEP-1998;      98US-0099596.
PR      09-SEP-1998;      98US-0099598.
PR      09-SEP-1998;      98US-0099602.
PR      10-SEP-1998;      98US-0099642.
PR      10-SEP-1998;      98US-0099741.
PR      10-SEP-1998;      98US-0099754.
PR      10-SEP-1998;      98US-0099763.
PR      10-SEP-1998;      98US-0099792.
PR      10-SEP-1998;      98US-0099808.
PR      10-SEP-1998;      98US-0099812.
PR      10-SEP-1998;      98US-0099815.
PR      10-SEP-1998;      98US-0099816.
PR      15-SEP-1998;      98US-0100385.
PR      15-SEP-1998;      98US-0100388.
PR      15-SEP-1998;      98US-0100390.
PR      16-SEP-1998;      98US-0100584.
PR      16-SEP-1998;      98US-0100627.
PR      16-SEP-1998;      98US-0100661.
PR      16-SEP-1998;      98US-0100662.
PR      16-SEP-1998;      98US-0100664.
PR      17-SEP-1998;      98US-0100683.
PR      17-SEP-1998;      98US-0100684.
PR      17-SEP-1998;      98US-0100710.
PR      17-SEP-1998;      98US-0100711.
PR      17-SEP-1998;      98US-0100919.
PR      17-SEP-1998;      98US-0100930.
PR      18-SEP-1998;      98US-0100848.
PR      18-SEP-1998;      98US-0100849.
PR      18-SEP-1998;      98US-0100849.
PR      18-SEP-1998;      98US-0101014.
PR      18-SEP-1998;      98US-0101068.
PR      18-SEP-1998;      98US-0101071.
PR      22-SEP-1998;      98US-0101279.
PR      23-SEP-1998;      98US-0101471.
PR      23-SEP-1998;      98US-0101472.
PR      23-SEP-1998;      98US-0101474.
PR      23-SEP-1998;      98US-0101474.
PR      23-SEP-1998;      98US-0101475.
PR      23-SEP-1998;      98US-0101476.
PR      23-SEP-1998;      98US-0101477.
PR      23-SEP-1998;      98US-0101479.
PR      24-SEP-1998;      98US-0101738.
PR      24-SEP-1998;      98US-0101741.
PR      24-SEP-1998;      98US-0101743.
PR      24-SEP-1998;      98US-0101915.
PR      24-SEP-1998;      98US-0101916.
PR      29-SEP-1998;      98US-0102207.
PR      29-SEP-1998;      98US-0102240.
PR      29-SEP-1998;      98US-0102307.
PR      29-SEP-1998;      98US-0102330.
PR      29-SEP-1998;      98US-0102331.
PR      30-SEP-1998;      98US-0102484.
PR      30-SEP-1998;      98US-0102487.
PR      30-SEP-1998;      98US-0102570.
PR      30-SEP-1998;      98US-0102571.
PR      01-OCT-1998;      98US-0102684.

PR      01-OCT-1998;      98US-0102687.
PR      02-OCT-1998;      98US-0102965.
PR      06-OCT-1998;      98US-0103258.
PR      06-OCT-1998;      98US-0103449.
PR      07-OCT-1998;      98US-0103314.
PR      07-OCT-1998;      98US-0103315.
PR      07-OCT-1998;      98US-0103328.
PR      07-OCT-1998;      98US-0103395.
PR      07-OCT-1998;      98US-0103396.
PR      07-OCT-1998;      98US-0103401.
PR      08-OCT-1998;      98US-0103633.
PR      08-OCT-1998;      98US-0103678.
PR      08-OCT-1998;      98US-0103679.
PR      08-OCT-1998;      98US-0103711.
PR      14-OCT-1998;      98US-0104257.
PR      20-OCT-1998;      98US-0104987.
PR      20-OCT-1998;      98US-0105000.
PR      20-OCT-1998;      98US-0105002.
PR      21-OCT-1998;      98US-0105104.
PR      22-OCT-1998;      98US-0105169.
PR      22-OCT-1998;      98US-0105266.
PR      26-OCT-1998;      98US-0105693.
PR      26-OCT-1998;      98US-0105694.
PR      26-OCT-1998;      98US-0105807.
PR      27-OCT-1998;      98US-0105881.
PR      27-OCT-1998;      98US-0105882.
PR      27-OCT-1998;      98US-0106062.
PR      28-OCT-1998;      98US-0106023.
PR      28-OCT-1998;      98US-0106029.
PR      28-OCT-1998;      98US-0106030.
PR      28-OCT-1998;      98US-0106032.
PR      28-OCT-1998;      98US-0106033.
PR      28-OCT-1998;      98US-0106178.
PR      29-OCT-1998;      98US-0106248.
PR      29-OCT-1998;      98US-0106384.
PR      29-OCT-1998;      98US-0106500.
PR      30-OCT-1998;      98US-0106464.
PR      30-NOV-1998;      98US-0106856.
PR      03-NOV-1998;      98US-0106902.
PR      03-NOV-1998;      98US-0106905.
PR      03-NOV-1998;      98US-0106919.
PR      03-NOV-1998;      98US-0106932.
PR      03-NOV-1998;      98US-0106934.
PR      10-NOV-1998;      98US-0107783.
PR      17-NOV-1998;      98US-0108775.
PR      17-NOV-1998;      98US-0108779.
PR      17-NOV-1998;      98US-0108787.
PR      17-NOV-1998;      98US-0108788.
PR      17-NOV-1998;      98US-0108801.
PR      17-NOV-1998;      98US-0108802.
PR      17-NOV-1998;      98US-0108806.
PR      17-NOV-1998;      98US-0108807.
PR      17-NOV-1998;      98US-0108867.
PR      17-NOV-1998;      98US-0108925.
PR      18-NOV-1998;      98US-0108848.
PR      18-NOV-1998;      98US-0108849.
PR      18-NOV-1998;      98US-0108850.
PR      18-NOV-1998;      98US-0108851.
PR      18-NOV-1998;      98US-0108852.
PR      18-NOV-1998;      98US-0108858.
PR      18-NOV-1998;      98US-0108904.

XX      (GETH ) GENENTECH INC.
XX      Baker K, Goddard A, Gurney AL, Smith V, Watanabe CX, Wood WI;
XX      WPI, 2000-237871/20.
XX      N-PDB: AAA37116.
XX      New mammalian DNA sequences encoding transmembrane, receptor or
XX      secreted PRO polypeptides, useful for screening of potential peptide or
XX      small molecule inhibitors of the relevant receptor/ligand interactions
```

Query Match 85.7%; Score 30; DB 21; Length 215;
 Best Local Similarity 62.5%; Pred. No. 65;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxafxg 8
 11 11 1
 Db 25 wkvtaf1g 32

RESULT 11

AAB64401
 ID AAB64401 standard; Protein; 217 AA.

AC AAB64401;

DT 22-MAR-2001 (first entry)

DE Amino acid sequence of human Intracellular signalling molecule INTRA33.

XX Human; Intracellular signalling molecule; INTRA; Immunosuppressive;
 KW cytosolic; neuroprotective; neurotropic; antiarteriosclerotic; cancer;
 KW antiinflammatory; anti-HIV; neuroleptic; antibacterial; antifungal;
 KW antiviral; antiparasitic; antihelminthic; antiparkinsonian; AIDS;
 KW cell proliferative disorder; arteriosclerosis autoimmune; epilepsy;
 KW inflammatory disorder; Addison's disease; gastrointestinal disorder;
 KW neurological disorder; Parkinson's disease; Creutzfeldt-Jakob disease;
 KW mental disorder; schizophrenia; anxiety.

XX Homo sapiens.

PN WO200077040-A2.

PD 21-DEC-2000.

PE 16-JUN-2000; 2000WO-US16636.

PR 16-JUN-1999; 99US-0139566.

PR 17-AUG-1999; 99US-0149640.

PR 09-NOV-1999; 99US-0164417.

XX (INCYTE GENOMICS INC.

PA Yue H, Tang YT, Hillman JL, Lal P, Bandman O, Baughn MR;

PI Azimzal Y, Yang J, Reddy R, Lu DAM;

XX WPI: 2001-025334/03.

DR N-PSDB: AAF32670.

PT New human intracellular signalling molecules, useful for the diagnosis,

PT prevention and treatment of cell proliferative, autoimmune,

PT inflammatory, neurological, gastrointestinal, reproductive and

PT developmental disorders -

XX Claim 5; Page 141-142; 192pp; English.

PS Sequences AAF32638 - AAF32689 represent cDNA encoding human

CC intracellular signalling molecules INTRA1 - INTRA52, represented in

CC AAB64369 - AAB64420. Modulators of the intracellular signalling molecules

CC of the invention exhibit immunosuppressive; cytosolic; neuroprotective;

CC neurotropic; antiarteriosclerotic; antiinflammatory; anti-HIV;

CC neuroleptic; antibacterial; antifungal; antiviral; antiparasitic;

CC antihelminthic; and antiparkinsonian activity. INTRA polypeptides their

CC agonists and antagonists are useful for the treatment of a condition

CC associated with decreased or increased expression of functional INTRA.

CC Disorders associated with abnormal INTRA expression or activity include

CC cell proliferative disorders e.g. arteriosclerosis and cancers;

CC autoimmune or inflammatory disorders e.g. Addison's disease and acquired

CC immunodeficiency syndrome (AIDS); viral, bacterial, fungal, parasitic,

CC pyrozoal and helminthic infections; gastrointestinal disorders e.g.

CC dysphagia and irritable bowel syndrome; neurological disorders e.g.

CC epilepsy and Parkinson's disease; prion diseases e.g. Creutzfeldt-Jakob

CC disease and mental disorders e.g. anxiety, schizophrenia and Tourette's

CC disorder. Antibodies immuno specific for the INTRA proteins may also be

CC useful in the diagnosis of the above disorders.
 XX SQ Sequence 217 AA;

Query Match 85.7%; Score 30; DB 22; Length 217;
 Best Local Similarity 62.5%; Pred. No. 66;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxafxg 8
 11 11 1
 Db 30 wkvtaf1g 37

RESULT 12

AAV51681
 ID AAV51681 standard; Protein; 219 AA.

AC AAV51681;

DT 02-JUN-2000 (first entry)

DE Murine clodin 6 protein.

XX Clodin 6; murine; tight junction-constituting membrane protein;
 KW medicine.

XX Mus sp.

PN JP2000032984-A.

PD 02-FEB-2000.

PE 26-JUN-1998; 98JP-0179847.

PR 15-MAY-1998; 98JP-0133215.

PA (EISA) EISAI CO LTD.

DR WPI: 2000-285512/25.

DR N-PSDB: AAZ89153.

PT Tight junction-constituting membrane protein clodin family - useful in

PT the medical field

XX Claim 5; Page 17-18; 22pp; Japanese.

CC This invention describes novel murine nucleic acid sequences encoding the

CC clodin family of tight junction (TJ)-constituting membrane protein. The

CC membrane protein can be used in medical field. This sequence represents

CC the clodin 6 protein described in the method of the invention.

XX Sequence 219 AA;

Query Match 85.7%; Score 30; DB 21; Length 219;
 Best Local Similarity 62.5%; Pred. No. 66;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxafxg 8
 11 11 1
 Db 30 wkvtaf1g 37

RESULT 13
 AAY99434
 ID AAY99434 standard; Protein; 220 AA.

AC AAY99434;

DT 08-AUG-2000 (first entry)

DE Human PRO1488 (UNQ757) amino acid sequence SEQ ID NO:330.

Query Match 85.7%; Score 30; DB 21; Length 10;
Best Local Similarity 62.5%; Pred. No. 3.8;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8
| | | | |
Db 2 wkvtafig 9

RESULT 9

AAB24453
ID AAB24453 standard; Protein; 126 AA.

AC AAB24453;

DT 20-NOV-2000 (first entry)

DE Human secreted protein sequence encoded by gene 17 SEQ ID NO:78.

XX Human; secreted protein; cytosolic; antianaemic; antidiabetic;
KW antiinflammatory; ophthalmological; antirheumatic; antiarthritic;
KW antipsoriatic; antiangiogenic; cardiant; anti-HIV; neutropic;
KW neuroprotective; antimicrobial; antiparkinsonian; cancer;
KW immune system disorder; angiogenesis; hyperproliferative disorder;
KW cardiovascular disorder; apoptosis; neurological disease;
XX infectious disease; wound healing.

XX Homo sapiens.

OS WO200035937-A1.

PN 22-JUN-2000.

PD 16-DEC-1999; 99WO-US29950.

PF 17-DEC-1998; 98US-0112809.

PR 18-DEC-1998; 98US-0113006.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Ebner R, Rosen CA, Endress GA, Soppet DR, Ni J;
PI Duan DR, Moore PA, Shi Y, Lafleur DW, Olsen HS, Florence K;
XX N-PSDB; AAA78397.

DR WPI: 2000-431566/37.

PT Forty seven human nucleic acids encoding secreted proteins, useful in
PT the treatment, prevention and diagnosis of cancers, disorders of the
PT immune system, angiogenesis disorders, neurological diseases and
PT hyperproliferative disorders -
XX
XX
PS Claim 11; Page 492; 562pp; English.

XX The polynucleotide sequence given in AAA78381 to AAA78432 encode the
CC human secreted proteins given in AAB24437 to AAB24604. Human secreted
CC proteins have activities based on the tissues and cells the genes are
CC expressed in. Examples of activities include: cytosolic; antianaemic;
CC antidiabetic; antiinflammatory; ophthalmological; antirheumatic;
CC antiarthritic; antipsoriatic; antiangiogenic; cardiant; anti-HIV;
CC neutropic; neuroprotective; antimicrobial and antiparkinsonian.
CC Human secreted protein polynucleotides, polypeptides, antagonists and/or
CC agonists may be useful in treating, preventing, and/or diagnosing other
CC diseases, disorders, and/or conditions such as: (a) cancers; (b)
CC disorders of the immune system; (c) angiogenesis disorders; (d)
CC hyperproliferative disorders; (e) cardiovascular disorders; (f) diseases
CC associated with increase apoptosis; (g) neurological diseases; and
CC (h) infectious diseases. They are also used to promote wound healing.
CC AAA78372 to AAA78380 and AAB24436 represent sequences used in the
CC exemplification of the present invention.

XX Sequence 126 AA;

Query Match 85.7%; Score 30; DB 21; Length 126;
Best Local Similarity 62.5%; Pred. No. 40;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8
| | | | |
Db 25 wkvtafig 32

RESULT 10

AAB24485
ID AAB24485 standard; Protein; 215 AA.

AC AAB24485;

DT 20-NOV-2000 (first entry)

DE Human secreted protein sequence encoded by gene 17 SEQ ID NO:110.

XX Human; secreted protein; cytosolic; antianaemic; antidiabetic;
KW antiinflammatory; ophthalmological; antirheumatic; antiarthritic;
KW antipsoriatic; antiangiogenic; cardiant; anti-HIV; neutropic;
KW neuroprotective; antimicrobial; antiparkinsonian; cancer;
KW immune system disorder; angiogenesis; hyperproliferative disorder;
KW cardiovascular disorder; apoptosis; neurological disease;
XX infectious disease; wound healing.

XX Homo sapiens.

OS WO200035937-A1.

PN 22-JUN-2000.

PD 16-DEC-1999; 99WO-US29950.

PF 17-DEC-1998; 98US-0112809.

PR 18-DEC-1998; 98US-0113006.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Ebner R, Rosen CA, Endress GA, Soppet DR, Ni J;
PI Duan DR, Moore PA, Shi Y, Lafleur DW, Olsen HS, Florence K;
XX N-PSDB; AAA78429.

DR WPI: 2000-431566/37.

PT Forty seven human nucleic acids encoding secreted proteins, useful in
PT the treatment, prevention and diagnosis of cancers, disorders of the
PT immune system, angiogenesis disorders, neurological diseases and
PT hyperproliferative disorders -
XX
XX
PS Claim 11; Page 524-525; 562pp; English.

XX The polynucleotide sequence given in AAA78381 to AAA78432 encode the
CC human secreted proteins given in AAB24437 to AAB24604. Human secreted
CC proteins have activities based on the tissues and cells the genes are
CC expressed in. Examples of activities include: cytosolic; antianaemic;
CC antidiabetic; antiinflammatory; ophthalmological; antirheumatic;
CC antiarthritic; antipsoriatic; antiangiogenic; cardiant; anti-HIV;
CC neutropic; neuroprotective; antimicrobial and antiparkinsonian.
CC Human secreted protein polynucleotides, polypeptides, antagonists and/or
CC agonists may be useful in treating, preventing, and/or diagnosing other
CC diseases, disorders, and/or conditions such as: (a) cancers; (b)
CC disorders of the immune system; (c) angiogenesis disorders; (d)
CC hyperproliferative disorders; (e) cardiovascular disorders; (f) diseases
CC associated with increase apoptosis; (g) neurological diseases; and
CC (h) infectious diseases. They are also used to promote wound healing.
CC AAA78372 to AAA78380 and AAB24436 represent sequences used in the
CC exemplification of the present invention.

XX Sequence 215 AA;

PR 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
PA (ADHE-) ADHEREX TECHNOLOGIES INC.
PL Blaschuck OW, Symonds JM, Gour BJ;
DR WPI; 2000-365610/31.
XX
XX
XX Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
XX
PS Claim 73; Page 104; 121pp; English.
CC
CC The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-6 and claudin-9 groups of proteins
CC are cadherins, which are membrane glycoproteins involved in cell
CC adhesion. In some situations, cell adhesion occurs at abnormal levels,
CC and these peptides can be used to modulate these levels, and thus treat
CC autoimmune diseases, inflammatory diseases and cancer, and aid wound
CC healing and implant adhesion. In addition, they can also be used to
CC facilitate drug delivery to the desired target site. The present
CC sequence has a cyclic conformation.
XX
XX
SQ Sequence 10 AA;

Query Match 85.7%; Score 30; DB 21; Length 10;
Best Local Similarity 62.5%; Pred. No. 3.8;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8
11111
Db 2 wkvtatfg 9

RESULT 7
AAB06798
ID AAB06798 standard; peptide; 10 AA.
XX
AC AAB06798;
XX
DT 28-SEP-2000 (first entry)
XX
DE Claudin-6/9 cyclic cell adhesion recognition sequence SEQ ID NO: 369.
XX
XX
XX Claudin-6 modulating agent; claudin-9 modulating agent;
KW cell adhesion recognition sequence; CAR sequence; autoimmune disease;
KW inflammatory disease; cancer; graft rejection; cyclic.
XX
XX Mammalia.
OS
XX
XX WO200026360-A1.
PN
XX
XX 11-MAY-2000.
PD
XX
XX 03-NOV-1999; 99WO-CA01029.
PF
XX
XX 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
XX
XX (ADHE-) ADHEREX TECHNOLOGIES INC.
PA
PI Blaschuck OW, Symonds JM, Gour BJ;
XX
XX WPI; 2000-365610/31.
DR
XX
XX Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
XX
PS Claim 73; Page 104; 121pp; English.

XX
XX The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-6 and claudin-9 groups of proteins
CC are cadherins, which are membrane glycoproteins involved in cell
CC adhesion. In some situations, cell adhesion occurs at abnormal levels,
CC and these peptides can be used to modulate these levels, and thus treat
CC autoimmune diseases, inflammatory diseases and cancer, and aid wound
CC healing and implant adhesion. In addition, they can also be used to
CC facilitate drug delivery to the desired target site. The present
CC sequence has a cyclic conformation.
XX
XX
SQ Sequence 10 AA;

Query Match 85.7%; Score 30; DB 21; Length 10;
Best Local Similarity 62.5%; Pred. No. 3.8;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8
11111
Db 2 wkvtatfg 9

RESULT 8
AAB06806
ID AAB06806 standard; peptide; 10 AA.
XX
AC AAB06806;
XX
DT 28-SEP-2000 (first entry)
XX
DE Claudin-6/9 cyclic cell adhesion recognition sequence SEQ ID NO: 377.
XX
XX
XX Claudin-6 modulating agent; claudin-9 modulating agent;
KW cell adhesion recognition sequence; CAR sequence; autoimmune disease;
KW inflammatory disease; cancer; graft rejection; cyclic.
XX
XX Mammalia.
OS
XX
XX WO200026360-A1.
PN
XX
XX 11-MAY-2000.
PD
XX
XX 03-NOV-1999; 99WO-CA01029.
PF
XX
XX 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
XX
XX (ADHE-) ADHEREX TECHNOLOGIES INC.
PA
PI Blaschuck OW, Symonds JM, Gour BJ;
XX
XX WPI; 2000-365610/31.
DR
XX
XX Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
XX
PS Claim 73; Page 104; 121pp; English.
CC
CC The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-6 and claudin-9 groups of proteins
CC are cadherins, which are membrane glycoproteins involved in cell
CC adhesion. In some situations, cell adhesion occurs at abnormal levels,
CC and these peptides can be used to modulate these levels, and thus treat
CC autoimmune diseases, inflammatory diseases and cancer, and aid wound
CC healing and implant adhesion. In addition, they can also be used to
CC facilitate drug delivery to the desired target site. The present
CC sequence has a cyclic conformation.
XX
XX
SQ Sequence 10 AA;

OY 1 wkxxafxg 8
|| || |
Db 1 wkvtafig 8

RESULT 4

AAB06772
ID AAB06772 standard; peptide; 10 AA.

AC AAB06772;

DT 28-SEP-2000 (first entry)

DE Claudin-6/9 cyclic cell adhesion recognition sequence SEQ ID NO: 343.

XX Claudin-6 modulating agent; claudin-9 modulating agent;

KW cell adhesion recognition sequence; CAR sequence; autoimmune disease;

KM inflammatory disease; cancer; graft rejection; cyclic.

XX Mammalia.

PN WO200026360-A1.

XX 11-MAY-2000.

PF 03-NOV-1999; 99WO-CA01029.

PR 03-NOV-1998; 98US-0185908.

PR 30-MAR-1999; 99US-0282029.

PA (ADHE-) ADHEREX TECHNOLOGIES INC.

XX Blaschuck OW, Symonds JM, Gour BJ;

DR WPI; 2000-365610/31.

PT Antibody modulation of claudin-mediated cell adhesion for increasing

XX PT vasopermeability, for delivering drugs to tumors and the nervous system

XX PT and across the skin -

PS Claim 73; Page 103; 121pp; English.

XX The present invention relates to the use of peptides as claudin-mediated

CC cell adhesion modulators. The claudin-6 and claudin-9 groups of proteins

CC are cadherins, which are membrane glycoproteins involved in cell

CC adhesion. In some situations, cell adhesion occurs at abnormal levels,

CC and these peptides can be used to modulate these levels, and thus treat

CC autoimmune diseases, inflammatory diseases and cancer, and aid wound

CC healing and implant adhesion. In addition, they can also be used to

CC facilitate drug delivery to the desired target site. The present

CC sequence has a cyclic conformation.

XX Sequence 10 AA:

Query Match 85.7%; Score 30; DB 21; Length 10;

Best Local Similarity 62.5%; Pred. No. 3.8;

Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8

|| || |

Db 2 wkvtafig 9

|| || |

RESULT 5

AAB06781
ID AAB06781 standard; peptide; 10 AA.

AC AAB06781;

DT 28-SEP-2000 (first entry)

DE Claudin-6/9 cyclic cell adhesion recognition sequence SEQ ID NO: 352.

XX Claudin-6 modulating agent; claudin-9 modulating agent;

KW cell adhesion recognition sequence; CAR sequence; autoimmune disease;

KM inflammatory disease; cancer; graft rejection; cyclic.

XX Mammalia.

PN WO200026360-A1.

XX 11-MAY-2000.

PF 03-NOV-1999; 99WO-CA01029.

PR 03-NOV-1998; 98US-0185908.

PR 30-MAR-1999; 99US-0282029.

PA (ADHE-) ADHEREX TECHNOLOGIES INC.

XX Blaschuck OW, Symonds JM, Gour BJ;

DR WPI; 2000-365610/31.

PT Antibody modulation of claudin-mediated cell adhesion for increasing

XX PT vasopermeability, for delivering drugs to tumors and the nervous system

XX PT and across the skin -

PS Claim 73; Page 104; 121pp; English.

XX The present invention relates to the use of peptides as claudin-mediated

CC cell adhesion modulators. The claudin-6 and claudin-9 groups of proteins

CC are cadherins, which are membrane glycoproteins involved in cell

CC adhesion. In some situations, cell adhesion occurs at abnormal levels,

CC and these peptides can be used to modulate these levels, and thus treat

CC autoimmune diseases, inflammatory diseases and cancer, and aid wound

CC healing and implant adhesion. In addition, they can also be used to

CC facilitate drug delivery to the desired target site. The present

CC sequence has a cyclic conformation.

XX Sequence 10 AA;

Query Match 85.7%; Score 30; DB 21; Length 10;

Best Local Similarity 62.5%; Pred. No. 3.8;

Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8

|| || |

Db 2 wkvtafig 9

|| || |

RESULT 6

AAB06789
ID AAB06789 standard; peptide; 10 AA.

AC AAB06789;

DT 28-SEP-2000 (first entry)

DE Claudin-6/9 cyclic cell adhesion recognition sequence SEQ ID NO: 360.

XX Claudin-6 modulating agent; claudin-9 modulating agent;

KW cell adhesion recognition sequence; CAR sequence; autoimmune disease;

KM inflammatory disease; cancer; graft rejection; cyclic.

XX Mammalia.

PN WO200026360-A1.

XX 11-MAY-2000.

PF 03-NOV-1999; 99WO-CA01029.

DR WPI: 2000-365610/31.
XX Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
XX
PS Claim 70: Page 103; 121pp; English.
XX
CC The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-6 and claudin-9 groups of proteins
CC are cadherins, which are membrane glycoproteins involved in cell
CC adhesion. In some situations, cell adhesion occurs at abnormal levels,
CC and these peptides can be used to modulate these levels, and thus treat
CC autoimmune diseases, inflammatory diseases and cancer, and aid wound
CC healing and implant adhesion. In addition, they can also be used to
CC facilitate drug delivery to the desired target site.
XX
SQ Sequence 8 AA:

Query Match 85.7%; Score 30; DB 21; Length 8;
Best Local Similarity 62.5%; Pred. No. 4.3e+05;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxxxafxg 8
| | | | |
Db 1 wkvtafig 8

RESULT 2
AAB06814
ID AAB06814 standard; peptide; 8 AA.
XX
AC AAB06814:
XX
DT 28-SEP-2000 (first entry)
XX
DE Claudin-6/9 cyclic cell adhesion recognition sequence SEQ ID NO: 385.
XX
XX Claudin-6 modulating agent; claudin-9 modulating agent;
KW cell adhesion recognition sequence; CAR sequence; autoimmune disease;
KW inflammatory disease; cancer; graft rejection; cyclic.
XX
OS Mammalia.
XX
XX
XX M0200026360-A1.
XX
PD 11-MAY-2000.
XX
PE 03-NOV-1999; 99WO-CA01029.
XX
PR 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
XX
XX (ADHE-) ADHEREX TECHNOLOGIES INC.
XX
XX
XX Blaschuck OW, Symonds JM, Gour BJ;
PI WPI: 2000-365610/31.
XX
DR WPI: 2000-365610/31.
XX
PT Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
XX
PS Claim 73: Page 104; 121pp; English.
XX
CC The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-6 and claudin-9 groups of proteins
CC are cadherins, which are membrane glycoproteins involved in cell
CC adhesion. In some situations, cell adhesion occurs at abnormal levels,
CC and these peptides can be used to modulate these levels, and thus treat
CC autoimmune diseases, inflammatory diseases and cancer, and aid wound
CC healing and implant adhesion. In addition, they can also be used to

CC facilitate drug delivery to the desired target site. The present
CC sequence has a cyclic conformation.
XX
XX
SQ Sequence 8 AA:

Query Match 85.7%; Score 30; DB 21; Length 8;
Best Local Similarity 62.5%; Pred. No. 4.3e+05;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxxxafxg 8
| | | | |
Db 1 wkvtafig 8

RESULT 3
AAB06916
ID AAB06916 standard; protein; 8 AA.
XX
AC AAB06916:
XX
DT 05-OCT-2000 (first entry)
XX
DE Claudin cell adhesion recognition modulating sequence SEQ ID NO: 50.
XX
XX Claudin modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection.
XX
OS Synthetic.
XX
XX
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT FT Modified-site 8 /note= "C-terminal amide"
FT
XX
XX M0200026360-A1.
XX
PD 11-MAY-2000.
XX
PE 03-NOV-1999; 99WO-CA01029.
XX
PR 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
XX
XX (ADHE-) ADHEREX TECHNOLOGIES INC.
XX
XX
XX Blaschuck OW, Symonds JM, Gour BJ;
PI WPI: 2000-365610/31.
XX
DR WPI: 2000-365610/31.
XX
PT Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
XX
PS Claim 71: Page 103; 121pp; English.
XX
CC The present sequence is a peptide which can be used in a claudin-mediated
CC cell adhesion modulator. The claudin group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and this peptide
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, it can also be used to facilitate drug delivery
CC to the desired target site.
XX
XX
SQ Sequence 8 AA:

Query Match 85.7%; Score 30; DB 21; Length 8;
Best Local Similarity 62.5%; Pred. No. 4.3e+05;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL: AL512667; CAC21620.1; -;
DR InterPro: IPR001753; Enoyl-CoA_hydratase.
DR Pfam: PF00378; ECH; 1.
DR PROSITE: PS00166; ENOYL_COA_HYDRATASE; 1.
SQ SEQUENCE 266 AA; 28121 MW; C07F9346B82E0451 CRC64;

Query Match 80.0%; Score 28; DB 2; Length 266;
Best Local Similarity 50.0%; Pred. No. 96;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8
|: || |
Db 239 WRSVAFSG 246

RESULT 18

ID 09X982 PRELIMINARY; PRT; 311 AA.

AC 09X982;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE ORFAB.
OS ISRM10-1.
GN Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2011;
RT Martinez-Abarca F., Toro N.;
RL "Bacterial group II intron mobility into DNA transposition sites in vivo."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ242573; CAB43594.1; -;
SQ SEQUENCE 311 AA; 34583 MW; 90AF46A01EA80891 CRC64;

Query Match 80.0%; Score 28; DB 2; Length 311;
Best Local Similarity 50.0%; Pred. No. 11e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wkxxafxg 8
|: || |
Db 177 WKTTFGTG 184

RESULT 19

ID 09NME0 PRELIMINARY; PRT; 643 AA.

AC 09NME0;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHEtical. 73.3 KDA PROTEIN.
OS Homo sapiens (Human).
GN Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO;
RA Iisogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Aita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y., Oshima A.;
RT "NDO human cDNA sequencing project.";

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK000956; BAA91442.1; -;
DR InterPro: IPR000504; RRM.
DR Pfam: PF00076; rrm; 1.
DR SMART: SM00360; RRM; 1.
DR PROSITE: PS50102; RRM; 1.
SQ SEQUENCE 643 AA; 73288 MW; B6C7654AF7537CBA CRC64;

Query Match 80.0%; Score 28; DB 4; Length 643;
Best Local Similarity 62.5%; Pred. No. 2.4e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8
|: || |
Db 632 WKLCAPFG 639

RESULT 20

ID 09Z2L5 PRELIMINARY; PRT; 65 AA.

AC 09Z2L5;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE TRANSKETOLASE (FRAGMENT).
OS Mus musculus (Mouse).
GN Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129SV;
RL Salamon C., Sax C.M., Platiorsky J.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF046917; AAC80281.1; -;
DR HSSP: P23254; TRK.
DR InterPro: IPR000360; Transketolase.
DR Pfam: PF00456; transketolase; 1.
FT NON_TER 1 1
FT NON_TER 65 65
SQ SEQUENCE 65 AA; 7363 MW; B4E7FACFD2FA7BA8 CRC64;

Query Match 77.1%; Score 27; DB 11; Length 65;
Best Local Similarity 50.0%; Pred. No. 37;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8
|: || |
Db 18 WEAMAFAG 25

Search completed: January 14, 2002, 07:39:44
Job time: 955 sec

OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
Fuji F., Hiruma C., Nakamura Y., Ogasawara N., Kohara S.,
Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL: AP001512; BAB05321.1; -;
DR InterPro: IPR002912; ACT.
DR InterPro: IPR002162; D_2_hydroxyacid_DH.
DR Pfam: PF00389; 2-Hacd_DH; 1.
DR Pfam: PF01842; ACT; 1.
DR PROSITE: PS00670; D_2-HYDROXYACID_DH_2; 1.
DR PROSITE: PS00671; D_2-HYDROXYACID_DH_3; 1.
KM Complete proteome.
SQ SEQUENCE 540 AA; 59041 MW; C9D4A3AC5E49C35A CRC64;

Query Match 82.9%; Score 29; DB 2; Length 540;
Best Local Similarity 62.5%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8
DB 141 WKRAFOG 148

RESULT 15
09JHCO PRELIMINARY; PRT; 197 AA.
ID 09JHCO
AC 09JHCO;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE CBLN3.
GN CBLN3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Pang Z., Morgan J.I.;
RT "Cloning and characterization of a novel precerebellin-related gene."
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF218380; AAF32315.1; -;
DR EMBL: AF218379; AAF32314.1; -;
DR MGD: MGI:1889286; Cbln3.
DR InterPro: IPR001073; C1q.
DR Pfam: PF00386; C1q; 1.
DR PRINTS: PR00007; COMPLEMENTC1Q.
DR SMART: SM00110; C1Q; 1.
DR PROSITE: PS0113; C1Q; 1.
SQ SEQUENCE 197 AA; 21077 MW; DBA8925C9BB1B77 CRC64;

Query Match 80.0%; Score 28; DB 11; Length 197;
Best Local Similarity 50.0%; Pred. No. 70;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8
DB 184 WKSSFSG 191

RESULT 16
O9ADDO PRELIMINARY; PRT; 243 AA.
ID 09ADDO
AC 09ADDO;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 27.3 KDA PROTEIN.
GN SCBA5H2.06C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,
Kinashi H., Hopwood D.A.;

Query Match 80.0%; Score 28; DB 2; Length 243;
Best Local Similarity 50.0%; Pred. No. 87;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8
DB 47 WKASFLG 54

RESULT 17
09EWM0 PRELIMINARY; PRT; 266 AA.
ID 09EWM0
AC 09EWM0;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE PUTATIVE ENOYL-COA HYDRATASE.
GN 2SCR31.11C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,
Kinashi H., Hopwood D.A.;

```

OC Specmatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Malpighiales; Linaceae; Linum.
XX NCBI_TaxID=4006;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. MCGREGOR;
RA Jain R.K., Thompson R.G., Taylor D.C., Mackenzie S.L., McHughen A.G.,
RA Rowland G.G., Tenaschuk D., Coffey M.;
RT "Isolation of the two flax stearyl-acyl carrier protein desaturase
RT gene promoters by the inverse polymerase chain reaction and their
RT differential regulation in transgenic flax, tobacco, and canola.";
RL Submitted (May-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CONVERTS STEARYL-ACP TO OLEOYL-ACP BY INTRODUCTION OF A
CC CIS DOUBLE BOND BETWEEN CARBONS DELTA-9 AND DELTA-10 OF THE ACYL
CC CHAIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: STEARYL-[ACYL-CARRIER PROTEIN] + AH(2) + O(2)
CC = OLEOYL-[ACYL-CARRIER PROTEIN] + A + 2 H(2)O.
CC -1- COFACTOR: FERREDOXIN, FERREDOXIN NADPH REDUCTASE, AND NADPH (BY
CC SIMILARITY).
CC -1- PATHWAY: CATALYZES THE PRINCIPAL CONVERSION OF SATURATED FATTY
CC ACIDS TO UNSATURATED FATTY ACIDS IN THE SYNTHESIS OF VEGETABLE
CC OILS.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST (BY SIMILARITY).
CC -1- SIMILARITY: TO OTHER PLANTS STEARYL-ACP DESATURASE AND TO
CC CYANOBACTERIA FATTY ACID DESATURASE (DESA).
DR EMBL: AJ006958; CAA07350.1; -.
DR HSSP: P22337; IAFR.
DR InterPro: IPR001225; FA_desaturase.
DR Pfam: PF00487; FA_desaturase; 1.
DR PROSITE: PS00574; FATTY_ACID_DESATUR_2; 1.
DR Chloroplast; Fatty acid biosynthesis; NADP; Oxidoreductase;
KW Translt peptide.
SQ SEQUENCE 396 AA; 44881 MW; 565B42B12B0CFA23 CRC64;

Oy 1 wkxxafxg 8
    |||||
Db 340 WKVDAFTG 347

Query Match      85.7%; Score 30; DB 10; Length 396;
Best Local Similarity 62.5%; Pred. No. 51;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

RESULT 12
ID 097363 PRELIMINARY; PRT; 313 AA.
AC 097363;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE LIPOPOLYSACCHARIDE BINDING PROTEIN PRECURSOR.
OS Bombyx mori (Silk moth).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Bombycoidea; Bombycidae; Bombyx.
OX NCBI_TaxID=7091;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEMOCYTE;
RX MEDLINE=99142698; PubMed=98989592;
RA Koizumi N., Imanura M., Kadotani T., Yaoi K., Iwahana H., Sato R.;
RT "The lipopolysaccharide-binding protein participating in hemocyte
RT nodule formation in the silkworm Bombyx mori is a novel member of the
RT C-type lectin superfamily with two different tandem carbohydrate-
RT recognition domains.";
RL FEBS Lett. 443:139-143(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=HEMOCYTE;
RX MEDLINE=97454306; PubMed=9310381;
RA Koizumi N., Morozumi A., Imanura M., Tanaka E., Iwahana H., Sato R.;

```

```

RT "Lipopolysaccharide-binding proteins and their involvement in the
RT bacterial clearance from the hemolymph of the silkworm Bombyx mori.";
RL Eur. J. Biochem. 248:217-224(1997).
DR EMBL: AJ011573; CAB38429.1; -.
DR HSSP: P20693; IHLJ.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF00059; lectin_c; 2.
DR SMART: SM00034; CLECT; 2.
DR PROSITE: PS00615; C_TYPE_LLECTIN_1; UNKNOWN_1.
DR PROSITE: PS50041; C_TYPE_LLECTIN_2; 2.
KW Signal.
FT SIGNAL 1
FT CHAIN 20 19 POTENTIAL.
FT SEQUENCE 313 AA; 35203 MW; 3C31534D4538C5B8 CRC64;

Oy 1 wkxxafxg 8
    |||||
Db 230 WKVDATFG 237

Query Match      82.9%; Score 29; DB 5; Length 313;
Best Local Similarity 62.5%; Pred. No. 67;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

RESULT 13
ID 09NBV9 PRELIMINARY; PRT; 327 AA.
AC 09NBV9;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE IMMOLECTIN-2.
OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Sphingioidea; Sphingidae; Sphinginae; Manduca.
OX NCBI_TaxID=7130;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FAF BODY;
RA Yu X., Kanost M.R.;
RT "Immulectin-2, a lipopolysaccharide-specific lectin from an insect,
RT Manduca sexta, is induced in response to gram-negative bacteria.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF242202; AAF91316.1; -.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF00059; lectin_c; 2.
DR SMART: SM00034; CLECT; 2.
DR PROSITE: PS00615; C_TYPE_LLECTIN_1; UNKNOWN_1.
DR PROSITE: PS50041; C_TYPE_LLECTIN_2; 2.
SQ SEQUENCE 327 AA; 37423 MW; 3B7FB3E62C78AFB CRC64;

Query Match      82.9%; Score 29; DB 5; Length 327;
Best Local Similarity 62.5%; Pred. No. 71;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 wkxxafxg 8
    |||||
Db 232 WKVDATFG 239

RESULT 14
ID 09KCG9 PRELIMINARY; PRT; 540 AA.
AC 09KCG9;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE D-3-PHOSPHOGLYCERATE DEHYDROGENASE.
GN BH1602.
OS Bacillus halodurans.

```

RA Tetteijn H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Elsen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwin M.L., Deboy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T.D., Ciecko A., Parksey D.S., Blair E., Citrono H., Clark E.B.,
RA Cotton M.D., Ulteback T.R., Khouri H., Qin H., Yamachyan J.,
RA Gill J., Scarlato V., Maignani V., Pizsa M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.,
RT "Complete genome sequence of *Neisseria meningitidis* serogroup B strain
MC58."
RL Science 287:1809-1815(2000).
CC -1- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).
CC -1- SIMILARITY: TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE FAMILY.
DR EMBL: AL162756; CAB84746.1; -
DR EMBL: AE002479; AAF41679.1; -
DR TIGR: NMB304; -
DR InterPro: IPR002085; Adh_zinc.
DR InterPro: IPR002328; Adh_zinc.
DR Pfam: PF00107; adh_zinc; 1.
DR PROSITE: PS00059; Adh_ZINC; 1.
KW Complete proteome: Oxidoreductase; Zinc.
SQ SEQUENCE 378 AA; 40479 MW; 439PEE1606780D9A4 CRC64;

Query Match 85.7%; Score 30; DB 2; Length 378;
Best Local Similarity 62.5%; Pred. No. 49;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8
DB 318 WKSAFGG 325

RESULT 9
ID Q9FND2 PRELIMINARY; PRT: 379 AA.
AC Q9FND2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE ALCOHOL DEHYDROGENASE (EC 1.1.1.1) CLASS III.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID:3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:COLUMBIA;
RX MEDLINE:98069011; PubMed:9405937;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
Sequence features of the regions of 1,044,062 bp covered by thirteen
physically assigned P1 clones."
RL DNA Res. 4:291-300(1997).
CC -1- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).
CC EMBL: AB006703; BAB09054.1; -
DR InterPro: IPR002085; Adh_zinc.
DR InterPro: IPR002328; Adh_zinc.
DR InterPro: IPR002025; NAD_binding.
DR Pfam: PF00107; adh_zinc; 1.
DR PROSITE: PS00059; Adh_ZINC; 1.
KW Oxidoreductase; Zinc.
SQ SEQUENCE 379 AA; 40639 MW; 045054298F16B258 CRC64;

Query Match 85.7%; Score 30; DB 10; Length 379;
Best Local Similarity 62.5%; Pred. No. 49;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8
DB 14 111

DB 317 WKSAFGG 324

RESULT 10
ID 082014 PRELIMINARY; PRT: 396 AA.
AC 082014;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE ACYL-[ACYL-CARRIER PROTEIN] DESATURASE (EC 1.14.99.6) (STEAROYL-ACP
DESATURASE).
GN SAD1.
OS Linum usitatissimum (Flax) (Linseed).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Malpighiales; Linaceae; Linum.
OX NCBI_TaxID=4006;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:CV. MCGREGOR;
RA Jain R.K., Thompson R.G., Taylor D.C., Mackenzie S.L., McHughen A.G.,
RA Rowland G.G., Tenaschuk D., Coffey M.;
RT "Isolation of the two flax stearyl-acyl carrier protein desaturase
gene promoters by the inverse polymerase chain reaction and their
RT differential regulation in transgenic flax, tobacco, and canola."
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CONVERTS STEAROYL-ACP TO OLEOYL-ACP BY INTRODUCTION OF A
CC CIS DOUBLE BOND BETWEEN CARBONS DELTA-9 AND DELTA-10 OF THE ACYL
CC CHAIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: STEAROYL-[ACYL-CARRIER PROTEIN] + AH(2) + O(2)
CC = OLEOYL-[ACYL-CARRIER PROTEIN] + A + 2 H(2)O.
CC -1- COFACTOR: FERREDOXIN, FERREDOXIN NADPH REDUCTASE, AND NADPH (BY
CC SIMILARITY).
CC -1- PATHWAY: CATALYZES THE PRINCIPAL CONVERSION OF SATURATED FATTY
CC ACIDS TO UNSATURATED FATTY ACIDS IN THE SYNTHESIS OF VEGETABLE
CC OILS.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST (BY SIMILARITY).
CC -1- SIMILARITY: TO OTHER PLANTS STEAROYL-ACP DESATURASE AND TO
CC CYANOBACTERIA FATTY ACID DESATURASE (DESA).
DR EMBL: AJ006957; CA07349.1; -
DR HSSP: P22337; IAFR.
DR Mendel: 32621; Linus:1039;32621.
DR InterPro: IPR001225; FA_desaturase.
DR Pfam: PF00487; FA_desaturase; 1.
DR PROSITE: PS00574; FATTY_ACID_DESATUR_2; 1.
KW Chloroplast; Fatty acid biosynthesis; NADP; Oxidoreductase;
KW transit peptide.
SQ SEQUENCE 396 AA; 44857 MW; 938C1CDD273D827D CRC64;

Query Match 85.7%; Score 30; DB 10; Length 396;
Best Local Similarity 62.5%; Pred. No. 51;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8
DB 340 WKVDAFTG 347

RESULT 11
ID Q9SBA2 PRELIMINARY; PRT: 396 AA.
AC Q9SBA2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE ACYL-[ACYL-CARRIER PROTEIN] DESATURASE (EC 1.14.99.6) (STEAROYL-ACP
DESATURASE).
GN SAD2.
OS Linum usitatissimum (Flax) (Linseed).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

CC -1-COPFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).
DR EMBL: AF156708; AAF73255.1; -;
DR EMBL: AF156698; AAF73255.1; JOINED.
DR EMBL: AF156699; AAF73255.1; JOINED.
DR EMBL: AF156700; AAF73255.1; JOINED.
DR EMBL: AF156701; AAF73255.1; JOINED.
DR EMBL: AF156702; AAF73255.1; JOINED.
DR EMBL: AF156703; AAF73255.1; JOINED.
DR EMBL: AF156704; AAF73255.1; JOINED.
DR EMBL: AF156705; AAF73255.1; JOINED.
DR EMBL: AF156706; AAF73255.1; JOINED.
DR EMBL: AF156707; AAF73255.1; JOINED.
DR InterPro: IPR002085; Adh_zinc.
DR InterPro: IPR002328; Adh_zinc.
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00107; adh_zinc; 1.
DR PROSITE: PS00059; ADH_ZINC; 1.
DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
DR Oxidoreductase; zinc.
KW
RL SEQUENCE 377 AA; 40213 MW; C0F35E5208E438A1 CRC64;
SO

Query Match 85.7%; Score 30; DB 5; Length 377;
Best Local Similarity 62.5%; Pred. No. 48;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 wkxxafxg 8
DB 317 WKGTAFGG 324

RESULT 6
O9BJ34 PRELIMINARY; PRT; 377 AA.
AC O9BJ34;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ALCOHOL DEHYDROGENASE CLASS 3 (EC 1.1.1.1).
GN ADH3.
OS Branchiostoma floridae (Florida lancelet) (Amphioxus).
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma.
OX NCBI_TaxID=7739;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20507554; PubMed=11054102;
RA Canestro C., Hjelmgvist L., Albalat R., Garcia-Fernandez J.,
RA Gonzalez-Duarte R., Jornvall H.;
RT "Amphioxus alcohol dehydrogenase is a class 3 form of single type and
of structural conservation but with unique developmental expression.";
RL Eur. J. Biochem. 267:6511-6518(2000).
[2]
RP SEQUENCE FROM N.A.
RA Canestro C., Albalat R., Hjelmgvist L., Godoy L., Jornvall H.,
RA Gonzalez-Duarte R.;
RT "Ascidian and amphioxus Adh genes reveal functional and molecular
evolution of the ADH family expansion during vertebrate evolution.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF344170; AAK26851.1; -;
DR Oxidoreductase.
KW
RL SEQUENCE 377 AA; 40191 MW; 035FA2AD2890237B CRC64;
SO

Query Match 85.7%; Score 30; DB 5; Length 377;
Best Local Similarity 62.5%; Pred. No. 48;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 wkxxafxg 8
DB 317 WKGTAFGG 324

RESULT 7
O9BJ33 PRELIMINARY; PRT; 377 AA.
AC O9BJ33;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ALCOHOL DEHYDROGENASE CLASS 3 (EC 1.1.1.1).
GN ADH3.
OS Branchiostoma floridae (Florida lancelet) (Amphioxus).
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma.
OX NCBI_TaxID=7739;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20507554; PubMed=11054102;
RA Canestro C., Hjelmgvist L., Albalat R., Garcia-Fernandez J.,
RA Gonzalez-Duarte R., Jornvall H.;
RT "Amphioxus alcohol dehydrogenase is a class 3 form of single type and
evolution of the ADH family expansion during vertebrate evolution.";
RL Eur. J. Biochem. 267:6511-6518(2000).
[2]
RP SEQUENCE FROM N.A.
RA Canestro C., Albalat R., Hjelmgvist L., Godoy L., Jornvall H.,
RA Gonzalez-Duarte R.;
RT "Ascidian and amphioxus Adh genes reveal functional and molecular
evolution of the ADH family expansion during vertebrate evolution.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF344171; AAK26852.1; -;
DR Oxidoreductase.
KW
RL SEQUENCE 377 AA; 40104 MW; 707B83BA68371E10 CRC64;
SO

Query Match 85.7%; Score 30; DB 5; Length 377;
Best Local Similarity 62.5%; Pred. No. 48;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 wkxxafxg 8
DB 317 WKGTAFGG 324

RESULT 8
O9JRB0 PRELIMINARY; PRT; 378 AA.
AC O9JRB0;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ALCOHOL DEHYDROGENASE CLASS-III (EC 1.1.1.1) (ALCOHOL DEHYDROGENASE,
CLASS III).
GN ADHC OR NMA1518 OR NMB1304.
OS Neisseria meningitidis (serogroup A), and
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699; 491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis 22491.";
RL Nature 404:502-506(2000).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
SO

Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxfxg 8
DB 309 WKGSFAGC 316

RESULT 2

ID 024687 PRELIMINARY; PRT; 369 AA.

AC 024687;

DT 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE GLUTATHIONE DEPENDENT FORMALDEHYDE DEHYDROGENASE (EC 1.2.1.1).

GN GDFALDH.

OS Anabaena azollae.

OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.

OX NCBI_TaxID=1164;

RN [1]

RP SEQUENCE FROM N.A.

RA Shaw W.H., Muirhead A., Ariol T., Plazinski J.;

RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.

CC -1- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).

CC -1- SIMILARITY: TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE FAMILY.

DR EMBL: U69767; AAB1921.1; -.

DR HSSP: P11766; ITH.

DR InterPro: IPR002085; Adh_zinc.

DR InterPro: IPR002328; Adh_zinc.

DR InterPro: IPR000205; NAD_binding.

DR Pfam: PF00107; adh_zinc.1.

DR PROSITE: PS00059; ADH_ZINC: 1.

KM Oxidoreductase; Zinc.

SO SEQUENCE 369 AA; 39424 MW; 5E9F243D1024F594 CRC64;

QY 1 wxxxxfxg 8
DB 309 WKGSFAGC 316

Query Match
Best Local Similarity 85.7%; Score 30; DB 2; Length 369;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

RESULT 3

ID 09A5D4 PRELIMINARY; PRT; 369 AA.

AC 09A5D4;

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE ALCOHOL DEHYDROGENASE, CLASS III.

GN CC2516.

OS Caulobacter crescentus.

OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;

OC Caulobacter.

OX NCBI_TaxID=69394;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE-21173698; PubMed-11259647;

RA Nielsen W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,

RA Eلسen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,

RA Plocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,

RA Doboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,

RA Kolony J.F., Smit J., Craven M.B., Khouli H., Shetty J., Berry K.,

RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,

RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;

RT "Complete genome sequence of Caulobacter crescentus.";

RL Ploeg. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).

DR EMBL: AEO05919; AAK24487.1; -.

DR TIGR: CC2516; -.

KM Complete proteome.

SO SEQUENCE 369 AA; 39232 MW; 413C2B377BB41781 CRC64;

Query Match
Best Local Similarity 85.7%; Score 30; DB 2; Length 369;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxfxg 8
DB 309 WKGSFAGC 316

RESULT 4

ID 09NJD0 PRELIMINARY; PRT; 377 AA.

AC 09NJD0;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE ALCOHOL DEHYDROGENASE CLASS III (EC 1.1.1.1).

GN ADH.

OS Branchiostoma floridae (Florida lancelet) (Amphioxus).

OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;

OX NCBI_TaxID=7739;

RN [1]

RP SEQUENCE FROM N.A.

RA Canestro C., Hjelmqvist L., Albalat R., Garcia-Fernandez J.,

RA Gonzalez-Duarte R., Jornvall H.;

RT "Amphioxus ADH class III gene and protein characterization: family expansion and divergence times among major living forms.";

RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

CC -1- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).

DR EMBL: AF154331; AAF73254.1; -.

DR InterPro: IPR002085; Adh_zinc.

DR InterPro: IPR002328; Adh_zinc.

DR InterPro: IPR000531; TonB_boxc.

DR Pfam: PF00107; adh_zinc.1.

DR PROSITE: PS00059; ADH_ZINC: 1.

KM Oxidoreductase; Zinc.

SO SEQUENCE 377 AA; 40127 MW; 24FEF255DB242CCD CRC64;

QY 1 wxxxxfxg 8
DB 317 WKGSFAGC 324

Query Match
Best Local Similarity 85.7%; Score 30; DB 5; Length 377;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

RESULT 5

ID 09NJC3 PRELIMINARY; PRT; 377 AA.

AC 09NJC3;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE ALCOHOL DEHYDROGENASE CLASS III (EC 1.1.1.1).

GN ADH.

OS Branchiostoma lanceolatum (Common lancelet) (Amphioxus).

OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;

OX NCBI_TaxID=7740;

RN [1]

RP SEQUENCE FROM N.A.

RA Canestro C., Hjelmqvist L., Albalat R., Garcia-Fernandez J.,

RA Gonzalez-Duarte R., Jornvall H.;

RT "Amphioxus ADH class III gene and protein characterization: family expansion and divergence times among major living forms.";

RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:39:43 ; Search time 112.89 Seconds
(without alignments)
10.366 Million cell updates/sec

Title: 09-185908-1d
Perfect score: 35
Sequence: 1 wkxxafxg 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

- Database :
- 1: SPREMBL_17:*
 - 2: sp_archaea:*
 - 3: sp_bacteria:*
 - 4: sp_fungi:*
 - 5: sp_human:*
 - 6: sp_invertebrate:*
 - 7: sp_mammal:*
 - 8: sp_mhc:*
 - 9: sp_organelle:*
 - 10: sp_phage:*
 - 11: sp_plant:*
 - 12: sp_rodent:*
 - 13: sp_virus:*
 - 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	85.7	369	2	Q59399 escherichia
2	30	85.7	369	2	Q24687 anabaena az
3	30	85.7	369	2	Q9A5D4 caulobacter
4	30	85.7	377	5	Q9NJD0 branchiosto
5	30	85.7	377	5	Q9NJC3 branchiosto
6	30	85.7	377	5	Q9BJ34 branchiosto
7	30	85.7	377	5	Q9BJ33 branchiosto
8	30	85.7	378	2	Q9JRB0 neisseria m
9	30	85.7	379	10	Q9FND2 arabidopsis
10	30	85.7	396	10	Q82014 linum usita
11	30	85.7	396	10	Q9SBA2
12	29	82.9	313	5	Q97363 bombux mori
13	29	82.9	327	5	Q9KCV9 manduca sex
14	29	82.9	540	2	Q9KCG9 bacillus ha
15	28	80.0	197	11	Q9JHG0 mus musculi
16	28	80.0	243	2	Q9ADD0 streptomyce
17	28	80.0	266	2	Q9EWM0 streptomyce
18	28	80.0	311	2	Q9X982 rhizobium m
19	28	80.0	643	4	Q9NWE0 rhizobium saplen

ALIGNMENTS

20	27	77.1	65	11	Q9Z2L5	Q9Z2L5 mus musculu
21	27	77.1	158	4	Q9H667	Q9H667 homo sapien
22	27	77.1	159	9	Q38144	Q38144 bacterioph
23	27	77.1	214	13	Q9DE12	Q9DE12 xenopus lae
24	27	77.1	214	13	Q9S8R2	Q9S8R2 gallus gall
25	27	77.1	315	2	Q9X983	Q9X983 rhizobium m
26	27	77.1	315	2	Q9R9L5	Q9R9L5 rhizobium m
27	27	77.1	330	2	Q9S2V4	Q9S2V4 streptomyce
28	27	77.1	370	2	Q9HY01	Q9HY01 pseudomonas
29	27	77.1	372	2	Q9RD05	Q9RD05 pseudomonas
30	27	77.1	559	11	Q9RSA0	Q9RSA0 mus musculu
31	27	77.1	619	5	Q9VAK4	Q9VAK4 drosophila
32	27	77.1	895	10	Q9FGN6	Q9FGN6 arabidopsis
33	27	77.1	1639	27	Q9P707	Q9P707 schizosach
34	26	74.3	166	11	Q9GKK2	Q9GKK2 mus musculu
35	26	74.3	182	26	P97158	P97158 escherichia
36	26	74.3	208	13	Q9PSS0	Q9PSS0 gallus gall
37	26	74.3	229	9	Q9ZX28	Q9ZX28 mycobacteri
38	26	74.3	242	2	Q9I715	Q9I715 pseudomonas
39	26	74.3	247	2	Q9EXE4	Q9EXE4 escherichia
40	26	74.3	303	2	P97157	P97157 escherichia
41	26	74.3	316	1	Q26177	Q26177 methanobact
42	26	74.3	348	8	Q9G147	Q9G147 daphniaphyll
43	26	74.3	375	11	Q64533	Q64533 geomyx knox
44	26	74.3	375	11	Q64673	Q64673 geomyx knox
45	26	74.3	375	11	Q9Z2M3	Q9Z2M3 geomyx knox
46	26	74.3	375	11	Q9Z2M2	Q9Z2M2 geomyx knox
47	26	74.3	375	11	Q9Z2M1	Q9Z2M1 geomyx knox
48	26	74.3	375	11	Q9ROE4	Q9ROE4 geomyx knox
49	26	74.3	379	2	Q84731	Q84731 chlamydia t
50	26	74.3	379	2	Q9PLK0	Q9PLK0 chlamydia m

RESULT 1	Q59399	PRELIMINARY;	PRT;	369 AA.
ID	Q59399;			
AC	01-NOV-1996 (TEMBREL. 01, created)			
DT	01-NOV-1996 (TEMBREL. 01, last sequence update)			
DT	01-JUN-2001 (TEMBREL. 17, last annotation update)			
DE	FORMALDEHYDE DEHYDROGENASE (EC 1.2.1.1) (FORMALDEHYDE DEHYDROGENASE)			
DE	(GLUTATHIONE)) (FORMIC DEHYDROGENASE).			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Escherichia			
OX	NCBI_TaxID=562;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPRAIN-CLINICAL ISOLATE;			
RX	MEDLINE=97046207; PubMed=8891129;			
RA	Kremerle N., Feucht H., Kaulfers P.M.;			
RT	"Plasmid-mediated formaldehyde resistance in Escherichia coli:			
RT	Characterization of resistance gene";			
RL	Antimicrob. Agents Chemother. 40:2276-2279(1996).			
CC	-1- CATALYTIC ACTIVITY: FORMALDEHYDE + GLUTATHIONE + NAD(+) = S-			
CC	FORMYLGLUTATHIONE + NADH.			
CC	-1- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).			
CC	-1- SIMILARITY: TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE FAMILY.			
DR	EMBL: X73835; CA52057.1; -.			
DR	HSSP: P11766; ITIH.			
DR	InterPro: IPR002085; Adh_zinc.			
DR	InterPro: IPR002328; Adh_zinc.			
DR	Pfam: PF00107; adh_zinc; 1.			
DR	PROSITE, PS00059; ADH_ZINC; 1.			
KW	Oxidoreductase; Zinc.			
SQ	SEQUENCE 369 AA; 39089 MM; C7A1FF5EBF636C6D CRC64;			

Query Match 85.7%; Score 30; DB 2; Length 369;
Best Local Similarity 62.5%; Pred. No. 47;


```

RESULT 19
ADHX_RAT STANDARD: PRT: 373 AA.
ID ADHX_RAT STANDARD: PRT: 373 AA.
AC P12711;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ALCOHOL DEHYDROGENASE CLASS III (EC 1.1.1.1) (ALCOHOL DEHYDROGENASE
DE 2) (GLUTATHIONE-DEPENDENT FORMALDEHYDE DEHYDROGENASE) (EC 1.2.1.1)
DE (FDH) (FALDH) (ALCOHOL DEHYDROGENASE-B2).
GN ADH5 OR ADH2 OR ADH-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN RATTUS NORVEGICUS.
RP TISSUE: Liver;
RX MEDLINE=88152004; PubMed=3278908;
RA Julia P., Pares X., Joernvall H.;
RT "Rat liver alcohol dehydrogenase of class III. Primary structure,
RT functional consequences and relationships to other alcohol
RT dehydrogenases."
RL Eur. J. Biochem. 172:73-83(1988).
RN
RN
RP SEQUENCE OF 1-6.
RX MEDLINE=88005160; PubMed=3653405;
RA Falterwall T., Julia P., Kaiser R., Holmquist B., Pares X., Vallee B.L.,
RA Joernvall H.;
RT "Acetylated N-terminal structures of class III alcohol
RT dehydrogenases. Differences among the three enzyme classes."
RL FEBS Lett. 222:99-103(1987).
RN
RN
RP 3D-STRUCTURE MODELING.
RX MEDLINE=91363326; PubMed=1888714;
RA Lapatto R.;
RT "Model for the structure of formaldehyde dehydrogenase based on
RT Int. J. Biol. Macromol. 13:73-76(1991).
CC -1- FUNCTION: CLASS-III ADH IS REMARKABLY INEFFECTIVE IN OXIDIZING
CC ETHANOL, BUT IT READILY CATALYZES THE OXIDATION OF LONG-CHAIN
CC PRIMARY ALCOHOLS AND THE OXIDATION OF S-(HYDROXYMETHYL)
CC GLUTATHIONE.
CC -1- CATALYTIC ACTIVITY: ALCOHOL + NAD(+) = ALDEHYDE OR KETONE + NADH.
CC -1- CATALYTIC ACTIVITY: FORMALDEHYDE + GLUTATHIONE + NAD(+) =
CC S-FORMYLGLUTATHIONE + NADH.
CC -1- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY. CLASS-III SUBFAMILY.
DR HSSP; P11766; ITH.
DR InterPro: IPR002328; ADH_zinc.
DR InterPro: IPR002085; Adh_zinc.
DR Pfam: PF00107; adh_zinc; 1.
DR PROSITE; PS00059; ADH_ZINC; 1.
KW Oxidoreductase; Zinc; NAD; Multigene family; Acetylation.
FT MOD_RES 1
FT METAL 44
FT METAL 44
FT METAL 66
FT METAL 66
FT METAL 96
FT METAL 96
FT METAL 99
FT METAL 99
FT METAL 102
FT METAL 102
FT METAL 110
FT METAL 110
FT METAL 114
FT METAL 114
FT BINDING 114
FT BINDING 114
SQ SEQUENCE 373 AA: 39426 MW: 88084F8427D5A130 CRC64;

```

Query Match 85.7%; Score 30; DB 1; Length 373;
 Best Local Similarity 62.5%; Pred. No. 9.2;

Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxfxg 8
 DB 313 WKGTAFGC 320

```

RESULT 20
ADHX_UROHA STANDARD: PRT: 373 AA.
ID ADHX_UROHA STANDARD: PRT: 373 AA.
AC P80467;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ALCOHOL DEHYDROGENASE CLASS III (EC 1.1.1.1) (GLUTATHIONE-DEPENDENT
DE FORMALDEHYDE DEHYDROGENASE) (EC 1.2.1.1) (FDH).
OS Uromastix hardwickii (Indian spiny-tailed lizard).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Iguania; Acrodonta; Agamidae; Uromastycinae;
OC Uromastix.
OX NCBI_TaxID=40250;
RN
RN
RP SEQUENCE.
RX MEDLINE=96033975; PubMed=7589468;
RA Hjelmgvist L., Shafiqat J., Siddiqi A.R., Joernvall H.;
RT "Alcohol dehydrogenase of class III: consistent patterns of
RT structural and functional conservation in relation to class I and
RT other proteins."
RL FEBS Lett. 373:212-216(1995).
CC -1- FUNCTION: CLASS-III ADH IS REMARKABLY INEFFECTIVE IN OXIDIZING
CC ETHANOL, BUT IT READILY CATALYZES THE OXIDATION OF LONG-CHAIN
CC PRIMARY ALCOHOLS AND THE OXIDATION OF S-(HYDROXYMETHYL)
CC GLUTATHIONE.
CC -1- CATALYTIC ACTIVITY: ALCOHOL + NAD(+) = ALDEHYDE OR KETONE + NADH.
CC -1- CATALYTIC ACTIVITY: FORMALDEHYDE + GLUTATHIONE + NAD(+) =
CC S-FORMYLGLUTATHIONE + NADH.
CC -1- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY. CLASS-III SUBFAMILY.
DR HSSP; P11766; ITH.
DR InterPro: IPR002328; ADH_zinc.
DR InterPro: IPR002085; Adh_zinc.
DR Pfam: PF00107; adh_zinc; 1.
DR PROSITE; PS00059; ADH_ZINC; 1.
KW Oxidoreductase; Zinc; NAD; Multigene family; Acetylation.
FT MOD_RES 1
FT METAL 44
FT METAL 44
FT METAL 66
FT METAL 66
FT METAL 96
FT METAL 96
FT METAL 99
FT METAL 99
FT METAL 102
FT METAL 102
FT METAL 110
FT METAL 110
FT METAL 114
FT METAL 114
FT BINDING 114
FT BINDING 114
SQ SEQUENCE 373 AA: 39430 MW: 28B3B44A0A9FCF06 CRC64;

```

Query Match 85.7%; Score 30; DB 1; Length 373;
 Best Local Similarity 62.5%; Pred. No. 9.2;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxfxg 8
 DB 313 WKGTAFGC 320

Search completed: January 14, 2002, 07:40:38
 Job time: 504 sec

RX MEDLINE-93077045; PubMed-1446828;
 RA Hur M.W., Edenberg H.J.;
 RT "Cloning and characterization of the ADH5 gene encoding human alcohol
 RT dehydrogenase 5, formaldehyde dehydrogenase.";
 RL Gene 121:305-311(1992).
 RN [4]
 RP SEQUENCE.
 RC TISSUE=Liver;
 RA MEDLINE-88209465; PubMed-3365377;
 RX Kaler R., Holmquist B., Hempel J., Vallee B.L., Joernvall H.;
 RT "Class III human liver alcohol dehydrogenase: a novel structural type
 RT biochemically related to the class I and class II enzymes.";
 RL Biochemistry 27:1132-1140(1988).
 RN [5]
 RP PARTIAL SEQUENCE, AND MUTAGENESIS OF ARG-114.
 RA MEDLINE-93264427; PubMed-8494891;
 RX Holmquist B., Moulis J.M., Engeland K., Vallee B.L.;
 RT "Role of arginine 115 in fatty acid activation and formaldehyde
 RT dehydrogenase activity of human class III alcohol dehydrogenase.";
 RL Biochemistry 32:5139-5144(1993).
 RN [6]
 RP MUTAGENESIS OF ARG-114.
 RA MEDLINE-93211987; PubMed-8460164;
 RX Engeland K., Hoeoeg J.-O., Holmquist B., Estenius M., Joernvall H.,
 RA Vallee B.L.;
 RT "Mutation of Arg-115 of human class III alcohol dehydrogenase: a
 RT binding site required for formaldehyde dehydrogenase activity and
 RT fatty acid activation.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:2491-2494(1993).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
 RA MEDLINE-97170743; PubMed-9018047;
 RX Yang Z.-N., Boston W.F., Hurley T.D.;
 RT "Structure of human chl chl alcohol dehydrogenase: a glutathione-
 RT dependent formaldehyde dehydrogenase.";
 RL J. Mol. Biol. 265:330-343(1997).
 CC -1- FUNCTION: CLASS-III ADH IS REMARKABLY INEFFECTIVE IN OXIDIZING
 CC ETHANOL, BUT IT READILY CATALYZES THE OXIDATION OF LONG-CHAIN
 CC PRIMARY ALCOHOLS AND THE OXIDATION OF S-(HYDROXYMETHYL)
 CC GLUTATHIONE.
 CC -1- CATALYTIC ACTIVITY: ALCOHOL + NAD(+) = ALDEHYDE OR KETONE + NADH.
 CC -1- CATALYTIC ACTIVITY: FORMALDEHYDE + GLUTATHIONE + NAD(+) =
 CC S-FORMYLGLUTATHIONE + NADH.
 CC -1- COPACOPOR: REQUIRES ZINC FOR ITS ACTIVITY.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- MISCELLANEOUS: THERE ARE 7 DIFFERENT ADH'S ISOZYMES IN HUMAN:
 CC PI, ONE TO CLASS-I: ALPHA, BETA, AND GAMMA, ONE TO CLASS-II:
 CC CLASS-V: ADH6
 CC CLASS-III: CH1, ONE TO CLASS-IV: ADH7 AND ONE TO
 CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
 CC FAMILY. CLASS-III SUBFAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M30471: AAA79018.1: -
 DR EMBL: M29872: AAA51587.1: -
 DR EMBL: M81118: AAA51596.1: -
 DR EMBL: M81112: AAA51596.1: JOINED.
 DR EMBL: M81113: AAA51596.1: JOINED.
 DR EMBL: M81114: AAA51596.1: JOINED.
 DR EMBL: M81115: AAA51596.1: JOINED.
 DR EMBL: M81116: AAA51596.1: JOINED.
 DR EMBL: M81117: AAA51596.1: JOINED.
 DR PIR: A33428: DEHUC2.
 DR PIR: A36739: A36739.
 DR PIR: JH0789: JH0789.

DR PDB: 1TEH: 07-DEC-96.
 DR MIM: 103710: -
 DR MIM: 136490: -
 DR InterPro: IPR002328; ADH_zinc.
 DR InterPro: IPR002085; Adh_zinc.
 DR Pfam: PF00107; adh_zinc.1.
 DR PROSITE: PS00059; ADH_ZINC.1.
 KW Oxidoreductase; zinc; NAD; Multigene family; Acetylation;
 KW 3D-structure.
 FT INIT_MET 0 0
 FT MOD_RES 1 1
 FT METAL 44 44 ZINC (CATALYTIC).
 FT METAL 66 66 ZINC (CATALYTIC).
 FT METAL 96 96 ZINC (SECOND ATOM).
 FT METAL 99 99 ZINC (SECOND ATOM).
 FT METAL 102 102 ZINC (SECOND ATOM).
 FT METAL 110 110 ZINC (SECOND ATOM).
 FT METAL 173 173 ZINC (CATALYTIC).
 FT BINDING 114 114 IMPORTANT FOR FDH ACTIVITY AND ACTIVATION
 FT BY FATTY ACIDS.
 FT MUTAGEN 114 114 R->A,D: LOSS OF FDH ACTIVITY AND LOSS OF
 FT ACTIVATION BY FATTY ACIDS.
 FT CONFLICT 166 166 D -> Y (IN REF. 2).
 FT CONFLICT 245 245 F -> L (IN REF. 2).
 SQ SEQUENCE 373 AA; 39593 MW; A4EALABD6F0424F4 CRC64;

Query Match 85.7%; Score 30; DB 1; Length 373;
 Best Local Similarity 62.5%; Pred. No. 9.2;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 wkxxafxg 8
 Db 313 WKGTAFCG 320

RESULT 17
 ID ADHX_MOUSE STANDARD: PRT: 373 AA.
 AC P28474;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE ALCOHOL DEHYDROGENASE CLASS III (EC 1.1.1.1) (ALCOHOL DEHYDROGENASE
 DE 2) (GLUTATHIONE-DEPENDENT FORMALDEHYDE DEHYDROGENASE) (EC 1.2.1.1)
 DE (FDH) (FALDH) (ALCOHOL DEHYDROGENASE-B2) (ADH-B2).
 GN ADH5 OR ADH2 OR ADH-2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91272926; PubMed-2053480;
 RA Edenberg H.J., Brown C.J., Carr L.G., Ho W.H., Hur M.W.;
 RT "Alcohol dehydrogenase gene expression and cloning of the mouse
 RT chl-like ADH.";
 RL Adv. Exp. Med. Biol. 284:253-262(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-93112997; PubMed-1472709;
 RA Hur M.W., Ho W.H., Brown C.J., Goldman D., Edenberg H.J.;
 RT "Molecular cloning of mouse alcohol dehydrogenase-B2 cDNA: nucleotide
 RT sequences of the class III Adh genes evolve slowly even for silent
 RT substitutions";
 RL DNA Seq. 3:167-175(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SVJ;
 RX MEDLINE-96215448; PubMed-8647091;
 RA Foglio M.H., Duester G.;
 RT "Characterization of the functional gene encoding mouse class III
 RT alcohol dehydrogenase (glutathione-dependent formaldehyde

```

RL Unpublished observations (NOV-1994).
CC -1- CATALYTIC ACTIVITY: ALCOHOL + NAD(+) = ALDEHYDE OR KETONE + NADH.
CC -1- CATALYTIC ACTIVITY: FORMALDEHYDE + GLUTATHIONE + NAD(+) =
CC S-FORMYLGLUTATHIONE + NADH.
CC -1- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY, CLASS-III SUBFAMILY.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT HAD TO
CC BE INTRODUCED IN POSITION 128 TO PRODUCE THIS ORF.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: D16172; BAA03719.1; ALT_FRAME.
DR HSSP: P11766; 1TEH.
DR InterPro: IPR002328; ADH_zinc.
DR InterPro: IPR002085; Adh_zinc.
DR Pfam: PF00107; adh_zinc; 1.
DR PROSITE: PS00059; ADH_ZINC; 1.
KW Hypothetical protein; Oxidoreductase; Zinc; NAD; Plasmid.
FT METAL 40 40 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 42 42 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 92 92 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 95 95 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 98 98 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 106 106 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 169 169 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 369 AA; 39094 MW; ACNEM4581048901C CRC64;

Query Match 85.7%; Score 30; DB 1; Length 369;
Best Local Similarity 62.5%; Pred. No. 9.1;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxafxg 8
   || || |
Db 309 WKGTAFGG 316

RESULT 15
ADHX_HORSE STANDARD; PRT; 373 AA.
ID ADHX_HORSE
AC P19854;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ALCOHOL DEHYDROGENASE CLASS III CHAIN (EC 1.1.1.1) (GLUTATHIONE-
DE DEPENDENT FORMALDEHYDE DEHYDROGENASE) (EC 1.2.1.1) (FDH) (PALDH).
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RS SEQUENCE.
RX MEDLINE=90105360; PubMed=2690942;
RX Kaiser R., Holmquist B., Vallee B.L., Joernvall H.;
RX "Characteristics of mammalian class III alcohol dehydrogenases, an
RX enzyme less variable than the traditional liver enzyme of class I.";
RX Biochemistry 28:8432-8438(1989).
RL [2]
RS SEQUENCE OF 1-6.
RX MEDLINE=88005160; PubMed=3653405;
RX Fairwell T., Julia P., Kaiser R., Holmquist B., Pares X., Vallee B.L.,
RX Joernvall H.;
RX "Acetylated N-terminal structures of class III alcohol
RX dehydrogenases. Differences among the three enzyme classes.";

```

```

RL FEBS Lett. 222:99-103(1987).
CC -1- FUNCTION: CLASS-III ADH IS REMARKABLY INEFFECTIVE IN OXIDIZING
CC ETHANOL, BUT IT READILY CATALYZES THE OXIDATION OF LONG-CHAIN
CC PRIMARY ALCOHOLS AND THE OXIDATION OF S-(HYDROXYMETHYL)
CC GLUTATHIONE.
CC -1- CATALYTIC ACTIVITY: ALCOHOL + NAD(+) = ALDEHYDE OR KETONE + NADH.
CC -1- CATALYTIC ACTIVITY: FORMALDEHYDE + GLUTATHIONE + NAD(+) =
CC S-FORMYLGLUTATHIONE + NADH.
CC -1- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY, CLASS-III SUBFAMILY.
DR PIR: S02617; S02617.
DR PIR: A33419; A33419.
DR HSSP: P11766; 1TEH.
DR InterPro: IPR002328; ADH_zinc.
DR InterPro: IPR002085; Adh_zinc.
DR Pfam: PF00107; adh_zinc; 1.
DR PROSITE: PS00059; ADH_ZINC; 1.
KW Oxidoreductase; Zinc; NAD; Multigene family; Acetylation.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT METAL 44 44 ZINC (CATALYTIC).
FT METAL 66 66 ZINC (CATALYTIC).
FT METAL 96 96 ZINC (SECOND ATOM).
FT METAL 99 99 ZINC (SECOND ATOM).
FT METAL 102 102 ZINC (SECOND ATOM).
FT METAL 110 110 ZINC (SECOND ATOM).
FT METAL 173 173 ZINC (CATALYTIC).
FT BINDING 114 114 IMPORTANT FOR FDH ACTIVITY AND ACTIVATION
FT BY FATTY ACIDS (BY SIMILARITY).
SQ SEQUENCE 373 AA; 39440 MW; 491F01B0A7D3D32 CRC64;

Query Match 85.7%; Score 30; DB 1; Length 373;
Best Local Similarity 62.5%; Pred. No. 9.2;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxafxg 8
   || || |
Db 313 WKGTAFGG 320

RESULT 16
ADHX_HUMAN STANDARD; PRT; 373 AA.
ID ADHX_HUMAN
AC P11766;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ALCOHOL DEHYDROGENASE CLASS III CHAIN (EC 1.1.1.1) (GLUTATHIONE-
DE DEPENDENT FORMALDEHYDE DEHYDROGENASE) (EC 1.2.1.1) (FDH).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RS SEQUENCE FROM N.A.
RX MEDLINE=90056459; PubMed=2818582;
RX Sharma C.P., Fox E.A., Holmquist B., Joernvall H., Vallee B.L.;
RX "cDNA sequence of human class III alcohol dehydrogenase.";
RX Biochem. Biophys. Res. Commun. 164:631-637(1989).
RL [2]
RS SEQUENCE FROM N.A.
RX MEDLINE=90026418; PubMed=2679557;
RX Giri P.R., Krug J.F., Kozak C., Moretti T., O'Brien S.J.,
RX Senanez H.N., Goldman D.;
RX "Cloning and comparative mapping of a human class III (chi) alcohol
RX dehydrogenase cDNA.";
RX Biochem. Biophys. Res. Commun. 164:453-460(1989).
RL [3]
RS SEQUENCE FROM N.A.

```

KW Tlight junction: Transmembrane.
 FT TRANSMEM 8 28 POTENTIAL.
 FT TRANSMEM 82 102 POTENTIAL.
 FT TRANSMEM 117 137 POTENTIAL.
 FT TRANSMEM 161 181 POTENTIAL.
 SQ SEQUENCE 220 AA: 23277 MW: 6705065680DB909 CRC64:

Query Match
 Best Local Similarity 85.7%: Score 30; DB 1; Length 220;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxfxg 8
 DB 30 WKVAFIG 37

RESULT 13
 ADH3_ECOLI STANDARD; PRT; 369 AA.
 AC P25437; P75696; Q47533;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ALCOHOL DEHYDROGENASE CLASS III (EC 1.1.1.1) (GLUTATHIONE-DEPENDENT
 DE FORMALDEHYDE DEHYDROGENASE) (EC 1.2.1.1) (FADH) (FALDH).
 CN ADHC OR B0356.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OK NCBI_Taxid=562;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN K12;
 RA Nishimoto H., Saito N.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN 12
 RP SEQUENCE FROM N.A.
 RA Ito K., Matsumoto K., Tsuru D., Yoshimoto T.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 RN 13
 RP SEQUENCE FROM N.A.
 RC STRAIN K12 / MG1655;
 RX MEDLINE-97426617; PubMed-9278503;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN 14
 RP SEQUENCE FROM N.A.
 RA Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,
 RA Federle N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,
 RA Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN 15
 RP SEQUENCE OF 1-47.
 OS MEDLINE-92118844; PubMed-1731906;
 RX Guthell W.G., Holmquist B., Vallee B.L.;
 RT "Purification, characterization, and partial sequence of the
 RT glutathione-dependent formaldehyde dehydrogenase from Escherichia
 RT coli: a class III alcohol dehydrogenase.";
 RL Biochemistry 31:475-481(1992).
 CC -1- FUNCTION: HAS HIGH FORMALDEHYDE DEHYDROGENASE ACTIVITY IN THE
 CC PRESENCE OF GLUTATHIONE AND CATALYZES THE OXIDATION OF NORMAL
 CC ALCOHOLS IN A REACTION THAT IS NOT GSH-DEPENDENT. IN ADDITION,
 CC HEMITHIOLACETALS OTHER THAN THOSE FORMED FROM GSH, INCLUDING
 CC OMEGA-THIOL FATTY ACIDS, ALSO ARE SUBSTRATES.
 CC -1- CATALYTIC ACTIVITY: ALCOHOL + NAD(+) -> ALDEHYDE OR KETONE + NADH.
 CC -1- CATALYTIC ACTIVITY: FORMALDEHYDE + GLUTATHIONE + NAD(+) ->
 CC S-FORMYLGLUTATHIONE + NADH.
 CC -1- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (4 IONS OF ZINC PER

CC DIMER).
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
 CC FAMILY, CLASS-III SUBFAMILY.
 CC CAUTION: REF.1 AND REF.4 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
 CC FRAMESHIFT IN POSITION 26.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D85613; BAA12834.1; ALT_FRAME.
 DR EMBL: D38504; BAA22412.1; -.
 DR EMBL: AE000142; AAC73459.1; -.
 DR EMBL: U73857; AAB18081.1; ALT_FRAME.
 DR HSSP: P11766; ITH.
 DR EcoGene: EG50010; adhC.
 DR InterPro: IPR00328; ADH_zinc.
 DR InterPro: IPR002085; Adh_zinc.
 DR Pfam: PF00107; adh_zinc; 1.
 DR PROSITE: PS00059; ADH_ZINC; 1.
 KW Oxidoreductase; Zinc; NAD; Complete proteome.
 FT METAL 40 40 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 62 62 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 92 92 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 95 95 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 98 98 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 106 106 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 169 169 ZINC (CATALYTIC) (BY SIMILARITY).
 FT CONFLICT 41 41 H -> E (IN REF. 5).
 FT CONFLICT 46 46 T -> G (IN REF. 5).
 SQ SEQUENCE 369 AA; 39359 MW; 35B59078F8173521 CRC64;

Query Match
 Best Local Similarity 85.7%: Score 30; DB 1; Length 369;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxfxg 8
 DB 309 WKGSAPFG 316

RESULT 14
 ADH3_PASPI STANDARD; PRT; 369 AA.
 AC P39450;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE PUTATIVE ALCOHOL DEHYDROGENASE CLASS III (EC 1.1.1.1) (GLUTATHIONE-
 DE DEPENDENT FORMALDEHYDE DEHYDROGENASE) (EC 1.2.1.1) (FADH) (FALDH).
 OS Pasteurella piscicida (Photobacterium damsela (subsp. piscicida)).
 OG Plasmid.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae;
 OC Photobacterium.
 OK NCBI_Taxid=38294;
 RN 11
 RP SEQUENCE FROM N.A.
 RX MEDLINE-94328963; PubMed-8052160;
 RA Kim E.H., Aoki T.;
 RT "The transposon-like structure of IS26-tetracycline, and kanamycin
 RT resistance determinant derived from transferable R plasmid of fish
 RT pathogen, Pasteurella piscicida.";
 RL Microbiol. Immunol. 38:31-38(1994).
 RN 12
 RP IDENTIFICATION, AND CONCEPTUAL TRANSLATION.
 RA Rudd K.E., Balloch A.;

```

DT 30-MAY-2000 (Rel. 39, last sequence update)
DE 20-AUG-2001 (Rel. 40, last annotation update)
GN CLAUDIN-9.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN
RP SEQUENCE FROM N.A.
RA Morita K., Furuse M., Tsukita S.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AF124424; AAD17319.1; -.
DR MGI: 1913100; Cldn9.
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin; 1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN; 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 13 33 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 160 180 POTENTIAL.
SQ SEQUENCE 217 AA; 22984 MW; 2B46C7C43AE7EFD5 CRC64;

Query Match 85.7%; Score 30; DB 1; Length 217;
Best Local Similarity 62.5%; Pred. No. 5.5;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8
   || || |
DB 30 WKYTAFIG 37

RESULT 11
CLD6_MOUSE STANDARD; PRT; 219 AA.
AC 097262;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CLAUDIN-6.
GN CLDN6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN
RP SEQUENCE FROM N.A.
RA MEDLINE:99110921; PubMed:9892664;
RA Morita K., Furuse M., Fujimoto K., Tsukita S.;
RL "Claudin multigene family encoding four-transmembrane domain protein
RL components of tight junction strands."
RL Proc. Natl. Acad. Sci. U.S.A. 96:511-516(1999).
RN
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.

```

```

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AF087824; AAD09759.1; -.
DR EMBL: BC005718; AAH05718.1; -.
DR MGI: 1859284; Cldn6.
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR003925; Claudin6.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin; 1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN; 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 164 184 POTENTIAL.
SQ SEQUENCE 219 AA; 23387 MW; 00CF687FAB687E1F CRC64;

Query Match 85.7%; Score 30; DB 1; Length 219;
Best Local Similarity 62.5%; Pred. No. 5.6;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8
   || || |
DB 30 WKYTAFIG 37

RESULT 12
CLD6_HUMAN STANDARD; PRT; 220 AA.
AC P56747;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-6.
GN CLDN6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN
RP SEQUENCE FROM N.A.
RA Keen T.J., Inglehearn C.F.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AJ249735; CAB56533.1; -.
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR003925; Claudin6.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin; 1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN; 1.

```


CC PRIMARY ALCOHOLS AND THE OXIDATION OF S-(HYDROXYMETHYL)
CC GLUTATHIONE.
CC -1- CATALYTIC ACTIVITY: ALCOHOL + NAD(+) -> ALDEHYDE OR KETONE + NADH.
CC -1- CATALYTIC ACTIVITY: FORMALDEHYDE + GLUTATHIONE + NAD(+) ->
CC S-FORMYLGLUTATHIONE + NADH.
CC -1- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY.
CC -1- SUBUNIT: HOMODIMER AND HETERODIMER WITH L CHAIN.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY. CLASS-III SUBFAMILY.
CC InterPro: IPR002328; Adh_zinc.
DR InterPro: IPR002085; Adh_zinc.
DR Pfam: PF00107; adh_zinc; 1.
DR PROSITE: PS00059; ADH_ZINC; 1.
KW Oxidoreductase; Zinc; NAD; Multigene family; Acetylation.
FT MOD_RES 1 ACETYLATION.
FT METAL 45 46 ZINC (CATALYTIC).
FT METAL 68 68 ZINC (CATALYTIC).
FT METAL 98 98 ZINC (SECOND ATOM).
FT METAL 101 101 ZINC (SECOND ATOM).
FT METAL 104 104 ZINC (SECOND ATOM).
FT METAL 112 112 ZINC (SECOND ATOM).
FT METAL 175 175 ZINC (CATALYTIC).
FT BINDING 116 116 IMPORTANT FOR FDH ACTIVITY AND ACTIVATION
BY FATTY ACIDS (BY SIMILARITY).
SQ SEQUENCE 375 AA; 39669 MW; 0B9760AB7329FE3 CRC64;

Query Match 88.6%; Score 31; DB 1; Length 375;
Best Local Similarity 62.5%; Pred. No. 5.5;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxafxg 8
DB 315 WKATAFEG 322

RESULT 8
CLD9_BRNRE STANDARD; PRT; 208 AA.
AC 09YH91;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-LIKE PROTEIN /Z-F-89.
OS Brachydanio rerio (zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Keen T.J., Inglehearn C.F.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: AJ011789; CAA09777.1; -
DR InterPro: IPR001832; Claudin.
DR Pfam: PF00822; PMP22_Claudin.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN; 1.
KW Tight junction; Transmembrane.

FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 160 180 POTENTIAL.
SQ SEQUENCE 208 AA; 22205 MW; 6A4BB5BF3CCAB81 CRC64;

Query Match 85.7%; Score 30; DB 1; Length 208;
Best Local Similarity 62.5%; Pred. No. 5.3;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxafxg 8
DB 30 WKVTAFIG 37

RESULT 9
CLD9_HUMAN STANDARD; PRT; 217 AA.
AC 095484;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-9.
GN CLDN9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Keen T.J., Inglehearn C.F.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: AJ130941; CAA10254.1; -
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin; 1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN; 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 160 180 POTENTIAL.
SQ SEQUENCE 217 AA; 22848 MW; 0E49C5B5CB3AAC9C CRC64;

Query Match 85.7%; Score 30; DB 1; Length 217;
Best Local Similarity 62.5%; Pred. No. 5.3;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxafxg 8
DB 30 WKVTAFIG 37

RESULT 10
CLD9_MOUSE STANDARD; PRT; 217 AA.
AC 092057;
DT 30-MAY-2000 (Rel. 39, Created)

```
FT CHMIN 24 213 25 KDA OUTER-MEMBRANE IMMUNOGENIC
FT SEQUENCE 213 AA: 23185 MW: 59003CF46FD0ED3 CRC64:
SO
Query Match 88.6%; Score 31; DB 1; Length 213;
Best Local Similarity 62.5%; Pred. No. 3.2;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 wkxxafxg 8
111111
74 WKAGAFAG 81
Db
RESULT 5
OM25_BRUNE STANDARD; PRT; 213 AA.
AC 045326;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 25 KDA OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
GN OMP25.
OS Brucella neotomae.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
CC Brucellaceae; Brucella.
CX NCBI_TaxID=29460;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5K33;
RX MEDLINE=96239016; PubMed=8675306;
RA Cloeckert A., Verger J.M., Grayon M., Zygmunt M.S., Grepinet O.;
RT "Nucleotide sequence and expression of the gene encoding the major
RT 25-kilodalton outer membrane protein of Brucella ovis: Evidence for
RT antigenic shift, compared with other Brucella species, due to a
RT deletion in the gene."
RL Infect. Immun. 64:2047-2055(1996).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U39359; AAB36694.1; -.
CC InterPro: IPR000498; OmpA_tmem.
CC Pfam: PF01389; OmpA_membrane; 1.
CC Antigen: Outer membrane; Signal.
CC SIGNAL 1 23 POTENTIAL.
CC CHAIN 24 213 25 KDA OUTER-MEMBRANE IMMUNOGENIC
CC FT PROTEIN.
CC FT SEQUENCE 213 AA: 23167 MW: 49EE0F47B784F87 CRC64:
SO
Query Match 88.6%; Score 31; DB 1; Length 213;
Best Local Similarity 62.5%; Pred. No. 3.2;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 wkxxafxg 8
111111
74 WKAGAFAG 81
Db
RESULT 6
OM25_BRUSU STANDARD; PRT; 213 AA.
AC 045689;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
```

```
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 25 KDA OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
GN OMP25.
OS Brucella suis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
CC Brucellaceae; Brucella.
CX NCBI_TaxID=29461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1330.
RX MEDLINE=96239016; PubMed=8675306;
RA Cloeckert A., Verger J.M., Grayon M., Zygmunt M.S., Grepinet O.;
RT "Nucleotide sequence and expression of the gene encoding the major
RT 25-kilodalton outer membrane protein of Brucella ovis: Evidence for
RT antigenic shift, compared with other Brucella species, due to a
RT deletion in the gene."
RL Infect. Immun. 64:2047-2055(1996).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U39397; AAB36695.1; -.
CC InterPro: IPR000498; OmpA_tmem.
CC Pfam: PF01389; OmpA_membrane; 1.
CC Antigen: Outer membrane; Signal.
CC SIGNAL 1 23 POTENTIAL.
CC CHAIN 24 213 25 KDA OUTER-MEMBRANE IMMUNOGENIC
CC FT PROTEIN.
CC FT SEQUENCE 213 AA: 23151 MW: 55F33CF46FD0ED3 CRC64:
SO
Query Match 88.6%; Score 31; DB 1; Length 213;
Best Local Similarity 62.5%; Pred. No. 3.2;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 wkxxafxg 8
111111
74 WKAGAFAG 81
Db
RESULT 7
ADHH_GADMO STANDARD; PRT; 375 AA.
AC P81600;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ALCOHOL DEHYDROGENASE CLASS III H CHAIN (EC 1.2.1.1) (GDH).
DE DEPENDENT FORMALDEHYDE DEHYDROGENASE (EC 1.2.1.1) (FDH).
OS Gadus morhua (Atlantic cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae;
OC Gadus.
CX NCBI_TaxID=8049;
RN [1]
RP SEQUENCE.
RX MEDLINE=97085413; PubMed=8931553;
RA Danielsson O., Shafiqat J., Estonius M., El-Ahmad M., Joernvall H.;
RT "Isozyme multiplicity with anomalous dimer patterns in a class III
RT alcohol dehydrogenase. Effects on the activity and quaternary
RT structure of residue exchanges at 'non-functional' sites in a native
RT protein."
RL Biochemistry 35:14561-14568(1996).
CC -1- FUNCTION: CLASS-III ADH IS REMARKABLY INEFFECTIVE IN OXIDIZING
CC ETHANOL, BUT IT READILY CATALYZES THE OXIDATION OF LONG-CHAIN
```

```

DB      74 WKAGAFAG 81

RESULT  2
OM25_BRUB  STANDARD;          PRT: 213 AA.
AC      044664;
DT      15-DEC-1998 (Rel. 37, Created)
DT      15-DEC-1998 (Rel. 37, Last sequence update)
DE      25 KDA OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
GN      OMP25.
OS      Brucella abortus.
OC      Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC      Brucellaceae; Brucella.
OX      NCBI_TaxID:235;
RN      (1)
RP      SEQUENCE FROM N.A.
RC      STRAIN:5445;
RX      MEDLINE:95204367; PubMed:7896724;
RA      de Wergifosse P., Linermans P., Linet J.N., Cloeckaert A.;
RT      Cloning and nucleotide sequence of the gene coding for the major 25-
RL      kDa outer membrane protein of Brucella abortus."
CC      1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC      -1- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
CC      EMBL: X79284; CAAS5872.1; -.
DR      InterPro: IPR000498; OmpA_tmem.
DR      Pfam: PF01389; OmpA_membrane; 1.
KW      Antigen; Outer membrane; Signal.
FT      SIGNAL 1 23 POTENTIAL.
FT      CHAIN 24 213 25 KDA OUTER-MEMBRANE IMMUNOGENIC
FT      PROTEIN.
SQ      SEQUENCE 213 AA; 23052 MW; 2328515F1F794BC7 CRC64;

Query Match      88.6%; Score 31; DB 1; Length 213;
Best Local Similarity 62.5%; Pred. No. 3,2;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 wxxxxfxfg 8
DB      74 WKAGAFAG 81

RESULT  3
OM25_BRUCA  STANDARD;          PRT: 213 AA.
AC      045110;
DT      15-DEC-1998 (Rel. 37, Created)
DT      15-DEC-1998 (Rel. 37, Last sequence update)
DE      25 KDA OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
GN      OMP25.
OS      Brucella canis.
OC      Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC      Brucellaceae; Brucella.
OX      NCBI_TaxID:36855;
RN      (1)
RP      SEQUENCE FROM N.A.
RC      STRAIN:RM6/66;
RX      MEDLINE:96239016; PubMed:8675306;
RA      Cloeckaert A., Verger J.M., Grayon M., Zygmunt M.S., Grepinet O.;
RT      Nucleotide sequence and expression of the gene encoding the major
RT      "Nucleoside sequence and expression of the gene encoding the major

```

```

RT      25-Kilodalton outer membrane protein of Brucella ovis: Evidence for
RT      antigenic shift, compared with other Brucella species, due to a
RT      deletion in the gene."
RL      Infect. Immun. 64:2047-2055(1996).
CC      -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC      -1- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
CC      EMBL: U39358; AAB3692.1; -.
DR      InterPro: IPR000498; OmpA_tmem.
DR      Pfam: PF01389; OmpA_membrane; 1.
KW      Antigen; Outer membrane; Signal.
FT      SIGNAL 1 23 POTENTIAL.
FT      CHAIN 24 213 25 KDA OUTER-MEMBRANE IMMUNOGENIC
FT      PROTEIN.
SQ      SEQUENCE 213 AA; 23123 MW; D7E40E247A39B9DF CRC64;

Query Match      88.6%; Score 31; DB 1; Length 213;
Best Local Similarity 62.5%; Pred. No. 3,2;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 wxxxxfxfg 8
DB      74 WKAGAFAG 81

RESULT  4
OM25_BRUME  STANDARD;          PRT: 213 AA.
AC      045321;
DT      15-DEC-1998 (Rel. 37, Created)
DT      15-DEC-1998 (Rel. 37, Last sequence update)
DE      25 KDA OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
GN      OMP25.
OS      Brucella melitensis.
OC      Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC      Brucellaceae; Brucella.
OX      NCBI_TaxID:29459;
RN      (1)
RP      SEQUENCE FROM N.A.
RC      STRAIN:16M;
RX      MEDLINE:96239016; PubMed:8675306;
RA      Cloeckaert A., Verger J.M., Grayon M., Zygmunt M.S., Grepinet O.;
RT      Nucleotide sequence and expression of the gene encoding the major
RT      25-Kilodalton outer membrane protein of Brucella ovis: Evidence for
RT      antigenic shift, compared with other Brucella species, due to a
RT      deletion in the gene."
RL      Infect. Immun. 64:2047-2055(1996).
CC      -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC      -1- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
CC      EMBL: U33003; AAB06701.1; -.
DR      InterPro: IPR000498; OmpA_tmem.
DR      Pfam: PF01389; OmpA_membrane; 1.
KW      Antigen; Outer membrane; Signal.
FT      SIGNAL 1 23 POTENTIAL.

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:40:37 ; Search time 37.71 seconds
(without alignments)
7.778 Million cell updates/sec

Title: 09-185908-1D
Perfect score: 35
Sequence: 1 wkxxafxg 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : swissprot_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	31	88.6	201	1 OM25_BRUOV	Q45335 bruceella oy
2	31	88.6	213	1 OM25_BRUOB	Q44664 bruceella ab
3	31	88.6	213	1 OM25_BRUCA	Q45110 bruceella ca
4	31	88.6	213	1 OM25_BRUME	Q45321 bruceella me
5	31	88.6	213	1 OM25_BRUNE	Q45326 bruceella ne
6	31	88.6	213	1 OM25_BRUSU	Q45688 bruceella su
7	31	88.6	375	1 ADHX_GADMO	B81600 gadus morhu
8	30	85.7	208	1 CLD9_BRARE	Q3YH91 brachydanio
9	30	85.7	217	1 CLD9_HUMAN	Q3Y484 homo sapien
10	30	85.7	219	1 CLD9_MOUSE	Q92067 mus musculu
11	30	85.7	219	1 CLD6_MOUSE	Q92262 mus musculu
12	30	85.7	220	1 CLD6_HUMAN	P56747 homo sapien
13	30	85.7	369	1 ADH3_ECOLI	P25437 escherichia
14	30	85.7	369	1 ADH3_PASPI	P39450 pasteurella
15	30	85.7	373	1 ADHX_HORSE	P11984 equus cabal
16	30	85.7	373	1 ADHX_HUMAN	P11766 homo sapien
17	30	85.7	373	1 ADHX_MOUSE	P28474 mus musculu
18	30	85.7	373	1 ADHX_RABIT	O19053 oryctolagus
19	30	85.7	373	1 ADHX_RAT	P12711 rattus norv
20	30	85.7	373	1 ADHX_UROHA	P80467 uromastix h
21	30	85.7	375	1 ADHL_GADMO	P81601 gadus morhu
22	30	85.7	375	1 FADH_PARDE	P45382 paracoccus
23	30	85.7	376	1 ADHL_RHOSH	P72324 rhodobacter
24	30	85.7	376	1 ADHX_MYGGL	P80360 myxine glut
25	30	85.7	376	1 ADHX_SPRAU	P79896 sparus aupa
26	30	85.7	378	1 ADH3_HAEIN	P44557 haemophilus
27	30	85.7	378	1 ADHX_DROME	P46415 drosophila
28	30	85.7	378	1 ADHX_OCTVU	P81431 octopus vul
29	30	85.7	379	1 ADHX_PEA	P80572 pisum sativ
30	30	85.7	379	1 ADHX_ARATH	O96533 arabidopsis
31	30	85.7	381	1 ADHX_MAIZE	P93629 zea mays (m
32	30	85.7	381	1 ADHX_ORYSA	P93436 oryza sativ
33	30	85.7	381	1 FADH_CANMA	Q06099 candida mal

34	30	85.7	384	1 ADHX_CAEEL	Q17335 caenorhabdi
35	30	85.7	386	1 FADH_YEAST	P32771 saccharomyc
36	30	85.7	396	1 STAD_LINUS	P32062 linum usita
37	30	85.7	743	1 MYBH_XENLA	P52551 xenopus lae
38	29	82.9	199	1 VTAK_LAMBD	P03729 bacterioph
39	29	82.9	367	1 GCST_MYCLE	O32955 mycobacteri
40	29	82.9	662	1 LOXL_RAT	Q02759 rattus norv
41	29	82.9	662	1 LOXL_MOUSE	P39654 mus musculu
42	29	82.9	1220	1 DDH8_HUMAN	Q14562 homo sapien
43	28	80.0	528	1 SERA_MYCLE	O33116 mycobacteri
44	28	80.0	528	1 SERA_MYCTU	O53243 mycobacteri
45	27	77.1	193	1 CERB_HUMAN	P23435 homo sapien
46	27	77.1	193	1 CERB_MOUSE	Q97171 mus musculu
47	27	77.1	209	1 CLD4_CERAE	O19005 cercopithec
48	27	77.1	209	1 CLD4_HUMAN	O14493 homo sapien
49	27	77.1	209	1 CLD2_BRARE	Q3YH90 brachydanio
50	27	77.1	210	1 CLD4_MOUSE	O35054 mus musculu

ALIGNMENTS

```
RESULT 1
OM25_BRUOV STANDARD; PRT; 201 AA.
ID OM25_BRUOV
AC Q45335;
DT 15-DEC-1998 (Rel. 37, Created)
DI 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 25 KDA OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
GN OMP25.
OS Brucella ovis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=236;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=63/290;
RX MEDLINE=96239016; PubMed=8675306;
RA Cloeckaert A., Verger J.M., Grayon M., Zygmunt M.S., Grepinet O.;
RT "Nucleotide sequence and expression of the gene encoding the major
RT 25-kilodalton outer membrane protein of Brucella ovis: Evidence for
RT antigenic shift, compared with other Brucella species, due to a
RT deletion in the gene.";
RL Infect. Immun. 64:2047-2055(1996).
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY. HAS A C-TERMINAL
CC -!- DELETION COMPARED TO THAT OF OTHER BRUCELLA SPECIES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
CC EMBL: U33004; AAB06702.1; -.
DR InterPro: IPR000498; OmpA_tmem.
DR Pfam: PF01389; OmpA_membrane; 1.
KW Antigen; Outer membrane; Signal.
FT STGNAL 1 23 POTENTIAL.
FT CHAIN 24 201 25 KDA OUTER-MEMBRANE IMMUNOGENIC
FT PROTEIN.
SQ SEQUENCE 201 AA; 21817 MW; 85A4897489A0935B CRC64;
```

Query Match 88.6%; Score 31; DB 1; Length 201;
Best Local Similarity 62.5%; Pred. No. 3;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 wkxxafxg 8
|| || |

Mon Jan 14 08:24:48 2002

09-185908-1d.open.rpr

Page 8

Search completed: January 14, 2002, 07:37:28
Job time: 898 sec

C:Function: <ADH>
A:Description: catalyzes the oxidation of primary and secondary alcohols to aldehydes and
C:Function: <FDH>
A:Description: catalyzes the oxidation by NAD+ of formaldehyde and glutathione to S-form
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: alcohol metabolism; metalloprotein; NAD; oxidoreductase; zinc
F:34-370/Domain: long-chain alcohol dehydrogenase homology <LADH>
F:198-227/Region: beta-alpha-beta NAD nucleotide-binding fold
F:49,71,178/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
F:101,104,107,115/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match 85.7%; Score 30; DB 1; Length 381;
Best Local Similarity 62.5%; Pred. No. 23;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8
DB 320 WKGAFFG 327

RESULT 18
T03289
formaldehyde dehydrogenase (glutathione) (EC 1.2.1.1) - maize
C:Species: Zea mays (maize)
C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 08-Dec-2000
C:Accession: T03289
R:Fliegmann, J.; Sandermann, H.
Plant Mol. Biol. 34, 843-854, 1997
A:Title: Maize glutathione-dependent formaldehyde dehydrogenase cDNA: a novel plant gene
A:Reference number: Z14885; MUID:97435970
A:Accession: T03289
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-381 <FILE>
A:Cross-references: EMBL:Y11029; PIDN:CAA71913.1
A:Experimental source: cultivar Black Mexican Sweet
C:Genetics:
A:Note: fdh
C:Function: <GDH>
A:Description: catalyzes the NAD(+) - and glutathione-dependent oxidation of formaldehyde
A:Note: detoxifying enzyme
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: NAD; oxidoreductase; zinc
F:34-370/Domain: long-chain alcohol dehydrogenase homology <LADH>
F:49,71,179/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

Query Match 85.7%; Score 30; DB 2; Length 381;
Best Local Similarity 62.5%; Pred. No. 23;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8
DB 319 WKGAFFG 326

RESULT 19
T04164
formaldehyde dehydrogenase (glutathione) (EC 1.2.1.1) - rice
C:Species: Oryza sativa (rice)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 08-Dec-2000
C:Accession: T04164
R:Dolferus, R.; Osterman, J.C.; Peacock, W.J.; Dennis, E.S.
submitted to the EMBL Data Library, November 1996
A:Description: Cloning of the Arabidopsis and rice class III Adh genes: implications for
A:Reference number: Z15249
A:Accession: T04164
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-381 <DOI>
A:Cross-references: EMBL:U77637; NID:91675393; PIDN:AAB1911.1; PTD:91675394
A:Experimental source: subsp. Indica, var. IR36

C:Genetics:
A:Gene: AdhIII
A:Introns: 14/1; 59/3; 75/2; 184/1; 211/3; 237/1; 289/3; 343/3
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: NAD; oxidoreductase; zinc
F:34-370/Domain: long-chain alcohol dehydrogenase homology <LADH>
F:49,71,179/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

Query Match 85.7%; Score 30; DB 2; Length 381;
Best Local Similarity 62.5%; Pred. No. 23;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8
DB 319 WKGAFFG 326

RESULT 20
S31140
alcohol dehydrogenase (EC 1.1.1.1) SFA1 - yeast (Saccharomyces cerevisiae)
N:Alternate names: class III alcohol dehydrogenase; protein D1473; protein YDL168w
N:Contains: formaldehyde dehydrogenase (glutathione) (EC 1.2.1.1)
C:Species: Saccharomyces cerevisiae
C>Date: 28-May-1993 #sequence_revision 28-May-1993 #text_change 21-Jul-2000
C:Accession: S31140; S61045; S67720
R:Wehner, E.P.; Rao, E.; Brendel, M.
Mol. Gen. Genet. 237, 351-358, 1993
A:Title: Molecular structure and genetic regulation of SFA, a gene responsible for re
A:Reference number: S31138; MUID:93247548
A:Accession: S31140
A:Molecule type: DNA
A:Residues: 1-386 <MEH>
A:Cross-references: EMBL:X68020; NID:9577609; PIDN:CAA48161.1; PTD:9288591
R:Pohl, T.M.
submitted to the EMBL Data Library, November 1995
A:Reference number: S61010
A:Accession: S61045
A:Molecule type: DNA
A:Residues: 1-386 <POH>
A:Cross-references: EMBL:Z67750; NID:91061256; PIDN:CAA91578.1; PTD:91061271
R:Pohl, T.M.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S67708
A:Accession: S67720
A:Molecule type: DNA
A:Residues: 1-386 <POW>
A:Cross-references: EMBL:Z74216; NID:91431267; PIDN:CAA98742.1; PTD:91431268; GSPDB:G
C:Comment: Class III alcohol dehydrogenases are pyrazole-insensitive, are not very ac
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:SFA1; MIPS:YDL168w
A:Cross-references: SGD:S0002327; MIPS:YDL168w
A:Map position: 4L
C:Function: <ADH>
A:Description: catalyzes the oxidation of primary and secondary alcohols to aldehydes
C:Function: <FDH>
A:Description: catalyzes the oxidation by NAD+ of formaldehyde and glutathione to S-f
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: alcohol metabolism; metalloprotein; NAD; oxidoreductase; zinc
F:34-373/Domain: long-chain alcohol dehydrogenase homology <LADH>
F:199-228/Region: beta-alpha-beta NAD nucleotide-binding fold
F:49,71,179/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
F:101,104,107,115/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match 85.7%; Score 30; DB 1; Length 386;
Best Local Similarity 62.5%; Pred. No. 23;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8
DB 322 WKGAFFG 329

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Feb-2001
C:Accession: F81097.1 B81843
A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.
A:Reference number: AB1000; MUID:20175755
A:Accession: F81097
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-378 <RET>
A:Cross-references: GB:AE002479; GB:AE002098; NID:g7226543; PIDN:AAFA1679.1; PID:g7226543
A:Experimental source: serogroup B, strain MC58
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Moreau,
H.; Holtz, S.; Jorgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* 22491.
A:Reference number: AB1775; MUID:20222556
A:Accession: B81843
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-378 <PAR>
A:Cross-references: GB:AL162756; GB:AL157959; NID:g7380091; PIDN:CAB84746.1; PID:g7380161
A:Experimental source: serogroup A, strain 22491
C:Genetics:
A:Gene: adhC; NMB1304; NMA1518
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: NAD; oxidoreductase; zinc
F:49,771,178/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

Query Match 85.7%; Score 30; DB 2; Length 378;
Best Local Similarity 62.5%; Pred. No. 22;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 wxxaxfxg 8
|||
Db 318 WKSAFGC 325

RESULT 15
S51357
alcohol dehydrogenase (EC 1.1.1.1) Fdh - fruit fly (*Drosophila melanogaster*)
N:Alternate names: class III alcohol dehydrogenase
C:Species: *Drosophila melanogaster*
C:Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 03-Dec-1999
C:Accession: S51357
R:Luque, T.; Aclish, S.; Danielsson, O.; Joernvall, H.; Gonzalez-Duarte, R.
Eur. J. Biochem. 225, 985-993, 1994
A:Title: Structure of the *Drosophila melanogaster* glutathione-dependent formaldehyde de-
hydrogenase genes.
A:Reference number: S51357; MUID:95045575
A:Accession: S51357
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-379 <LUD>
A:Cross-references: EMBL:U07799; NID:g538264; PIDN:AA57187.1; PID:g538265
C:Comment: Class III alcohol dehydrogenases are pyrazole-insensitive, are not very active
C:Genetics:
A:Gene: FLYBae:Fdh
A:Cross-references: FlyBase:FBgn0011768
A:Introns: 7/3; 37/3
C:Complex: homodimer
C:Function: <ADH>
C:Function: <FDH>
A:Description: catalyzes the oxidation of primary and secondary alcohols to aldehydes and
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: acetylated amino end; alcohol metabolism; dimer; metalloprotein; NAD; oxidore-
ductase; 2-379/Product: alcohol dehydrogenase Fdh #status predicted <MAT>

```

F:33-370/Domain: long-chain alcohol dehydrogenase homology <LADH>
F:197-226/Region: beta-alpha-beta NAD nucleotide-binding fold
F:2/Modified site: acetylating amino end (Ser) (in mature form) #status predicted
F:48,70,177/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
F:100,103,106,114/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match          85.7%; Score 30; DB 1; Length 379;
Best Local Similarity 62.5%; Pred. No. 22;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 WXXXXFXG 8
          || || |
Db       319 WKGSAPFG 326

RESULT 16
S71244
alcohol dehydrogenase (EC 1.1.1.1) class III - Arabidopsis thaliana
N:Contains: formaldehyde dehydrogenase (glutathione) (EC 1.2.1.1)
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 28-Oct-1996 #sequence_revision 27-Feb-1997 #text_change 16-Jun-2000
C:Accession: S71244
R:Martinez, M.C.
Submitted to the EMBL Data Library, November 1994
A:Reference number: S71244
A:Accession: S71244
A:Molecule type: mRNA
A:Residues: 1-379 <NAR>
A:Cross-references: EMBL:X82647; NID:g1143387; PIDN:CAAS7973.1; PID:g1143388
C:Comment: Class III alcohol dehydrogenases are pyrazole-insensitive, are not very active
C:Function: <ADH>
A:Description: catalyzes the oxidation of primary and secondary alcohols to aldehydes
C:Function: <FDPH>
A:Description: catalyzes the oxidation of NAD+ of formaldehyde and glutathione to S-adenosylmethionine
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: alcohol metabolism; metalloprotein; NAD: oxidoreductase; zinc
F:32-368/Domain: long-chain alcohol dehydrogenase homology <LADH>
F:197-226/Region: beta-alpha-beta NAD nucleotide-binding fold
F:47,69,177/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
F:99,102,105,113/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match          85.7%; Score 30; DB 1; Length 379;
Best Local Similarity 62.5%; Pred. No. 22;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 WXXXXFXG 8
          || || |
Db       317 WKGTAFGG 324

RESULT 17
JN0447
alcohol dehydrogenase (EC 1.1.1.1) FDH1 - yeast (Candida maltosa)
N:Alternate names: Class III alcohol dehydrogenase
N:Contains: formaldehyde dehydrogenase (glutathione) (EC 1.2.1.1)
C:Species: Candida maltosa
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 11-Jun-1999
C:Accession: JN0447
R:Saenaukas, K.; Jomantien, R.; Januska, A.; Lebediene, E.; Lebedys, J.; Janulaitis
Gene 122, 207-211, 1992
A:Title: Cloning and analysis of a Candida maltosa gene which confers resistance to formaldehyde
A:Reference number: JN0447; MUID:93083986
A:Accession: JN0447
A:Molecule type: DNA
A:Residues: 1-381 <SAS>
A:Cross-references: GB:M58332; NID:g170870; PIDN:AAA34344.1; PID:g170871
C:Comment: This protein confers resistance to formaldehyde in yeast.
C:Comment: Class III alcohol dehydrogenases are pyrazole-insensitive, are not very active
C:Genetics:
A:Gene: FDH1
A:Introns: 8/3

```

Query Match 85.7%; Score 30; DB 1; Length 376;
Best Local Similarity 62.5%; Pred. No. 22;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8
|| || |
Db 316 WKGAFFG 323

RESULT 11
S51187
Alcohol dehydrogenase (EC 1.1.1.1) class III - Atlantic hagfish
N:Contains: formaldehyde dehydrogenase (glutathione) (EC 1.2.1.1)
C:Species: Myxine glutinosa (Atlantic hagfish)
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 26-Feb-1999
C:Accession: S51187; S66197
R:Danielsson, O.; Shafigat, J.; Estenius, M.; Joernvall, H.
Eur. J. Biochem. 225, 1081-1088, 1994
A:Title: Alcohol dehydrogenase class III contrasted to class I. Characterization of the
tion.
A:Reference number: S51187; MUID:95045537
A:Accession: S51187
A:Molecule type: Protein
A:Residues: 1-376 <DAN>
R:Jelmqvist, L.; Hackett, M.; Shafigat, J.; Danielsson, O.; Iida, J.; Hendrickson, R.C.;
FEBS Lett. 367, 237-240, 1995
A:Title: Multiplicity of N-terminal structures of medium-chain alcohol dehydrogenases. N
zyme.
A:Reference number: S66191; MUID:95331382
A:Accession: S66197
A:Molecule type: protein
A:Residues: 1-17 <HUE>
C:Comment: Class III alcohol dehydrogenases are pyrazole-insensitive, are not very activ
C:Complex: homodimer
C:Function: <ADH>
A:Description: catalyzes the oxidation of primary and secondary alcohols to aldehydes an
C:Function: <PDH>
A:Description: catalyzes the oxidation of primary and secondary alcohols to aldehydes an
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase; dimer; metalloprotein; NAD; oxidor
C:Keywords: acetylated amino end; alcohol metabolism; dimer; metalloprotein; NAD; oxidor
F:32-367/Domain: long-chain alcohol dehydrogenase homology <LADH>
F:196-225/Region: beta-alpha-beta NAD nucleotide-binding fold
F:1/Modified site: acetylated amino end (Ser) #status experimental
F:47,69,116/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
F:99,102,105,113/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match 85.7%; Score 30; DB 1; Length 376;
Best Local Similarity 62.5%; Pred. No. 22;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8
|| || |
Db 316 WKGAFFG 323

RESULT 12
A49662
alcohol dehydrogenase (EC 1.1.1.1) class III - common octopus
N:Contains: formaldehyde dehydrogenase (glutathione) (EC 1.2.1.1)
C:Species: Octopus vulgaris (common octopus)
C:Date: 07-Apr-1994 #sequence_revision 31-Dec-1995 #text_change 01-May-1998
C:Accession: A49662
R:Kaiser, R.; Fernandez, M.R.; Pares, X.; Joernvall, H.
Proc. Natl. Acad. Sci. U.S.A. 90, 11222-11226, 1993
A:Title: Origin of the human alcohol dehydrogenase system: implications from the structu
A:Reference number: A49662; MUID:94068576
A:Accession: A49662
A:Molecule type: Protein
A:Residues: 1-378 <KAI>
A:Experimental source: gills, salivary glands, heart

A:Note: sequence extracted from NCBI backbone (NCBI:140464)
C:Comment: Class III alcohol dehydrogenases are pyrazole-insensitive, are not very ac
C:Complex: homodimer
C:Function: <ADH>
A:Description: catalyzes the oxidation of primary and secondary alcohols to aldehydes
C:Function: <PDH>
A:Description: catalyzes the oxidation of primary and secondary alcohols to aldehydes
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase; dimer; metalloprotein; NAD; oxi
C:Keywords: acetylated amino end; alcohol metabolism; dimer; metalloprotein; NAD; oxi
F:1-378/Product: alcohol dehydrogenase class III #status experimental <MAT>
F:32-367/Domain: long-chain alcohol dehydrogenase homology <LADH>
F:196-225/Region: beta-alpha-beta NAD nucleotide-binding fold
F:1/Modified site: acetylated amino end (Thr) #status experimental
F:47,69,116/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
F:99,102,105,113/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match 85.7%; Score 30; DB 1; Length 378;
Best Local Similarity 62.5%; Pred. No. 22;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8
|| || |
Db 318 WKGAFFG 325

RESULT 13
H64052
alcohol dehydrogenase (EC 1.1.1.1) H10185 - Haemophilus influenzae (strain Rd KW20)
N:Alternate names: class III alcohol dehydrogenase
N:Contains: formaldehyde dehydrogenase (glutathione) (EC 1.2.1.1)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 11-Jun-1999
C:Accession: H64052
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630
A:Accession: H64052
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-378 <TRGR>
A:Cross-references: GB:032703; GB:I42023; NID:g1573133; PIDN:ANC21854.1; PID:g1573141
C:Comment: Class III alcohol dehydrogenases are pyrazole-insensitive, are not very ac
C:Complex: homodimer
C:Function: <ADH>
A:Description: catalyzes the oxidation of primary and secondary alcohols to aldehydes
C:Function: <PDH>
A:Description: catalyzes the oxidation of primary and secondary alcohols to aldehydes
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: alcohol metabolism; homodimer; metalloprotein; NAD; oxidoreductase; zinc
F:34-369/Domain: long-chain alcohol dehydrogenase homology <LADH>
F:198-227/Region: beta-alpha-beta NAD nucleotide-binding fold
F:49,71,118/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
F:101,104,107,115/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match 85.7%; Score 30; DB 1; Length 378;
Best Local Similarity 62.5%; Pred. No. 22;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxafxg 8
|| || |
Db 318 WKGAFFG 325

RESULT 14
F81097
probable alcohol dehydrogenase (EC 1.1.1.1) class III NMA1518 [similarity] - Neisseri
C:Species: Neisseria meningitidis

A:Reference number: A56397; MUID:95238401
A:Accession: A56397
A:Molecule type: DNA
A:Residues: 1-4 <H2>
A:Cross-references: GB:U010902; NID:9654723; PIDN:AAA83432.1; PID:91122906
R:Kaiser, R.; Holmquist, B.; Hempel, J.; Vallee, B.L.; Joernvall, H.
Biochemistry 27, 1132-1140, 1988
A:Title: Class III human liver alcohol dehydrogenase: a novel structural type equidistant
A:Reference number: A29983; MUID:88209465
A:Accession: A29983
A:Molecule type: protein
A:Residues: 2-374 <KA1>
R:Kaiser, R.; Holmquist, B.; Vallee, B.L.; Joernvall, H.
J. Protein Chem. 10, 69-73, 1991
A:Title: Human class III alcohol dehydrogenase/glutathione-dependent formaldehyde dehydratase
A:Reference number: A61428; MUID:91273763
A:Accession: A61428
A:Molecule type: protein
A:Residues: 2-374 <KA2>
R:Yang, Z.N.; Hurley, T.D.
submitted to the Brookhaven Protein Data Bank, February 1996
A:Reference number: A66716; PDB:1TEH
A:Contents: annotation: X-ray crystallography, 2.70 angstroms, residues 2-374
A:Comment: class III alcohol dehydrogenases are pyrazole-insensitive, are not very active
C:Genetics:
A:Gene: GDB:ADH5
A:Cross-references: GDB:118978; OMIM:103710
A:Map position: 4q21-4q25
A:Introns: 4/3; 38/3; 86/1; 115/2; 188/3; 275/3; 321/1; 367/2
A:Complex: homodimer; does not form heterodimers with the class I alcohol dehydrogenases
C:Function: <ADH>
A:Description: catalyzes the oxidation of primary and secondary alcohols to aldehydes and aldehydes to alcohols
A:Note: human alcohol dehydrogenase chl is expressed predominately in liver, kidney, and
C:Function: <FDB>
A:Description: catalyzes the oxidation by NAD+ of formaldehyde and glutathione to S-form
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: acetylated amino end; alcohol metabolism; homodimer; metalloprotein; NAD; ox
F:2-374/Product: alcohol dehydrogenase 5 #status experimental <MAT>
F:30-365/Domains: long-chain alcohol dehydrogenase homology <ADH>
F:194-223/Region: beta-alpha-beta NAD nucleotide-binding fold
F:2/Modified site: acetylated amino end (Ala) (in mature form) #status experimental
F:45,67,174/Binding site: zinc, catalytic (Cys, His, Cys) #status experimental
F:97,100,103,111/Binding site: zinc, noncatalytic (Cys) #status experimental

Query Match 85.7%; Score 30; DB 1; Length 374;
Best Local Similarity 62.5%; Pred. No. 22;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 wxxxafxg 8
DB 314 WKGRAFGG 321

RESULT 9
A56643
Alcohol dehydrogenase (EC 1.1.1.1) 2 - mouse
N:Alternate names: alcohol dehydrogenase chl chain; class III alcohol dehydrogenase
N:Contains: formaldehyde dehydrogenase (glutathione) (EC 1.2.1.1)
C:Species: Mus musculus (house mouse)
C:Date: 11-Aug-1995 #sequence, revision 11-Aug-1995 #text, change 11-Jun-1999
C:Accession: A56643; A60269; S71333
R:Hurt, M.W.; Ho, W.H.; Brown, C.J.; Goldman, D.; Edenberg, H.J.
DNA Seq. 3, 167-175, 1992
A:Title: Molecular cloning of mouse alcohol dehydrogenase-B2 cDNA: nucleotide sequences
A:Reference number: A56643; MUID:93112997
A:Accession: A56643
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-374 <HUR>
A:Cross-references: GB:M61417; GB:M61707; GB:M81315; NID:9191721; PIDN:AAA68896.1; PID:9
A:Note: sequence extracted from NCBI backbone (NCBIN:121466, NCBIPI:121468)
R:Edenberg, H.J.; Brown, C.J.; Carr, L.G.; Ho, W.H.; Hurt, M.W.

Adv. Exp. Med. Biol. 284, 253-262, 1991
A:Title: Alcohol dehydrogenase gene expression and cloning of the mouse-chl-like ADH.
A:Reference number: A60269; MUID:91272926
A:Accession: A60269
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 2-374 <ED6>
R:Foglio, M.H.; Duester, G.
Eur. J. Biochem. 237, 496-504, 1996
A:Title: Characterization of the functional gene encoding mouse class III alcohol dehy
ct open reading frame.
A:Reference number: S71333; MUID:96215448
A:Accession: S71333
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-374 <FOG>
A:Cross-references: EMBL:U48969; NID:91399475
A:Comment: Class III alcohol dehydrogenases are pyrazole-insensitive, are not very ac
C:Genetics:
A:Gene: adh-2
A:Introns: 4/3; 38/3; 86/1; 115/2; 188/3; 275/3; 321/1; 367/2
C:Complex: homodimer; does not form heterodimers with the class I alcohol dehydrogena
C:Function: <ADH>
A:Description: catalyzes the oxidation of primary and secondary alcohols to aldehydes
C:Function: <FDB>
A:Description: catalyzes the oxidation by NAD+ of formaldehyde and glutathione to S-f
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: acetylated amino end; alcohol metabolism; homodimer; metalloprotein; NAD;
F:2-374/Product: alcohol dehydrogenase 2 #status predicted <MAT>
F:30-365/Domains: long-chain alcohol dehydrogenase homology <ADH>
F:194-223/Region: beta-alpha-beta NAD nucleotide-binding fold
F:2/Modified site: acetylated amino end (Ala) (in mature form) #status predicted
F:45,67,174/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
F:97,100,103,111/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match 85.7%; Score 30; DB 1; Length 374;
Best Local Similarity 62.5%; Pred. No. 22;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 wxxxafxg 8
DB 314 WKGRAFGG 321

RESULT 10
JC4967
Alcohol dehydrogenase (EC 1.1.1.1) class III - gillhead sea bream
N:Contains: formaldehyde dehydrogenase (glutathione) (EC 1.2.1.1)
C:Species: Sparus aurata (gillhead sea bream)
C:Date: 31-Dec-1996 #sequence, revision 31-Dec-1996 #text, change 11-Jun-1999
C:Accession: JC4967
R:Funkenstein, B.; Jakowlew, S.B.
Gene 174, 159-164, 1996
A:Title: Molecular cloning of fish alcohol dehydrogenase cDNA.
A:Reference number: JC4967; MUID:97017142
A:Accession: JC4967
A:Molecule type: mRNA
A:Residues: 1-376 <RUN>
A:Cross-references: GB:U84791; NID:91814385; PIDN:AAA4188.1; PID:91814386
A:Comment: Class III alcohol dehydrogenases are pyrazole-insensitive, are not very ac
C:Complex: homodimer
C:Function: <ADH>
A:Description: catalyzes the oxidation of primary and secondary alcohols to aldehydes
C:Function: <FDB>
A:Description: catalyzes the oxidation by NAD+ of formaldehyde and glutathione to S-f
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: alcohol metabolism; dimer; metalloprotein; NAD; oxidoreductase; zinc
F:2-376/Product: alcohol dehydrogenase class III #status predicted <MAT>
F:32-367/Domains: long-chain alcohol dehydrogenase homology <ADH>
F:196-225/Region: beta-alpha-beta NAD nucleotide-binding fold
F:47,69,176/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
F:99,102,105,113/Binding site: zinc, noncatalytic (Cys) #status predicted

```

Db      313 WKGTAFGC 320

RESULT 6
DEPRTA
alcohol dehydrogenase (EC 1.1.1.1) 2 - rat
N:Alternate names: alcohol dehydrogenase chl chain; class III alcohol dehydrogenase
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 01-May-1998
C:Accession: S00331; S02619; S06633
R:Julia, P.; Pares, X.; Joernvall, H.
Eur. J. Biochem. 172, 73-83, 1988
A:Title: Rat liver alcohol dehydrogenase of class III. Primary structure, functional con
A:Reference number: S00331; MUID:88152004
A:Accession: S00331
A:Molecule type: protein
A:Residues: 1-373 <JUL>
R:Fairwell, T.; Julia, P.; Kaiser, R.; Holmquist, B.; Pares, X.; Vallee, B.L.; Joernvall
FEBS Lett. 222, 99-103, 1987
A:Title: Acetylated N-terminal structures of class III alcohol dehydrogenases. Differenc
A:Reference number: S02619; MUID:88005160
A:Accession: S02619
A:Molecule type: protein
A:Residues: 1-6 <FAI>
R:Kolusalo, M.; Baumann, M.; Uotila, L.
FEBS Lett. 257, 105-109, 1989
A:Title: Evidence for the identity of glutathione-dependent formaldehyde dehydrogenase
A:Reference number: S06633; MUID:90033321
A:Accession: S06633
A:Status: preliminary
A:Molecule type: protein
A:Residues: 9-25,84-95,'X',97-98,'X',187,188,189-193,'X',195-198;357-365 <KOI>
A:Experimental source: strain Wistar
C:Comment: Class III alcohol dehydrogenases are pyrazole-insensitive, are not very activ
C:Genetics:
A:Gene: ADH-2
A:Complex: homodimer; does not form heterodimers with the class I alcohol dehydrogenases
C:Function: <ADH>
A:Description: catalyzes the oxidation of primary and secondary alcohols to aldehydes an
C:Function: <FDB>
A:Description: catalyzes the oxidation by NAD+ of formaldehyde and glutathione to S-form
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: acetylated amino end; alcohol metabolism; homodimer; metalloprotein; NAD; ox
F:29-364/Product: alcohol dehydrogenase 2 #status experimental <MAT>
F:1-373/Product: alcohol dehydrogenase 2 #status experimental <LADH>
F:29-364/Domain: long-chain alcohol dehydrogenase homology
F:193-222/Region: beta-alpha-beta NAD nucleotide-binding fold
F:1/Modified site: acetylated amino end (Ala) #status experimental
F:44,66,173/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
F:96,99,102,110/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match      85.7%; Score 30; DB 1; Length 373;
Best local Similarity 62.5%; Pred. No. 22;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY      1 wkxxafgx 8
      || || |
Db      313 WKGTAFGC 320

RESULT 7
S68061
alcohol dehydrogenase (EC 1.1.1.1) class III - Indian spiny-tailed lizard
N:Contains: formaldehyde dehydrogenase (glutathione) (EC 1.2.1.1)
C:Species: Uromastyx hardwickii (Indian spiny-tailed lizard)
C:Date: 19-Mar-1997 #sequence_revision 29-Aug-1997 #text_change 01-May-1998
C:Accession: S68061; S66194
R:Hjelmqvist, L.; Shafqat, J.; Siddiqi, A.R.; Joernvall, H.
FEBS Lett. 373, 212-216, 1995
A:Title: Alcohol dehydrogenase of class III: consistent patterns of structural and funct
A:Reference number: S68061; MUID:96033975

```

```

A:Accession: S68061
A:Molecule type: protein
A:Residues: 1-373 <HJE>
R:Hjelmqvist, L.; Hackett, M.; Shafqat, J.; Danielsson, O.; Lida, J.; Hendrickson, R.
FEBS Lett. 367, 237-240, 1995
A:Title: Multiplicity of N-terminal structures of medium-chain alcohol dehydrogenases
nzyme.
A:Reference number: S66191; MUID:95331382
A:Accession: S66194
A:Molecule type: protein
A:Residues: 9-14 <HJM>
C:Comment: Class III alcohol dehydrogenases are pyrazole-insensitive, are not very ac
C:Complex: homodimer
C:Function: <ADH>
A:Description: catalyzes the oxidation of primary and secondary alcohols to aldehydes
C:Function: <FDB>
A:Description: catalyzes the oxidation by NAD+ of formaldehyde and glutathione to S-f
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: acetylated amino end; alcohol metabolism; dimer; metalloprotein; NAD; oxi
F:1-373/Product: alcohol dehydrogenase class III #status experimental <MAT>
F:29-364/Domain: long-chain alcohol dehydrogenase homology <LADH>
F:193-222/Region: beta-alpha-beta NAD nucleotide-binding fold
F:1/Modified site: acetylated amino end (Ala) #status experimental
F:44,66,173/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
F:96,99,102,110/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match      85.7%; Score 30; DB 1; Length 373;
Best local Similarity 62.5%; Pred. No. 22;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY      1 wkxxafgx 8
      || || |
Db      313 WKGTAFGC 320

RESULT 8
DEHUC2
alcohol dehydrogenase (EC 1.1.1.1) 5 [validated] - human
N:Alternate names: alcohol dehydrogenase chl chain; class III alcohol dehydrogenase
N:Contains: formaldehyde dehydrogenase (glutathione) (EC 1.2.1.1)
C:Species: Homo sapiens (man)
C:Date: 30-Sep-1991 #sequence_revision 05-Aug-1994 #text_change 15-Sep-2000
C:Accession: JH0789; A36739; A33428; A56397; A29983; A61428; S02618; S11077
R:Hur, M.W.; Edenberg, H.J.
Gene 121, 305-311, 1992
A:Title: Cloning and characterization of the ADH5 gene encoding human alcohol dehydro
A:Reference number: JH0789; MUID:93077045
A:Accession: JH0789
A:Molecule type: DNA
A:Residues: 1-374 <HUR>
A:Cross-references: GB:M81112; GB:M81118; NID:9178128; PIDN:AA51596.1; PID:9178130
A:Note: the authors present evidence that translation is initiated at Met-1
R:Sharma, C.P.; Fox, E.A.; Holmquist, B.; Joernvall, H.; Vallee, B.L.
Biochem. Biophys. Res. Commun. 164, 631-637, 1989
A:Title: CDNA sequence of human class III alcohol dehydrogenase.
A:Reference number: A36739; MUID:90056459
A:Accession: A36739
A:Molecule type: mRNA
A:Residues: 1-374 <SHA>
A:Cross-references: GB:M30471; NID:9178133; PIDN:AA49018.1; PID:9178134
R:Giri, P.R.; Krug, J.F.; Kozak, C.; Moretti, T.; O'Brien, S.J.; Seanez, H.N.; Goldm
Biochem. Biophys. Res. Commun. 164, 453-460, 1989
A:Title: Cloning and comparative mapping of a human class III (chl) alcohol dehydroge
A:Reference number: A33428; MUID:90026418
A:Accession: A33428
A:Molecule type: mRNA
A:Residues: MGAATPVDSPPRRPSVN',1-166,'Y',168-374 <GIR>
A:Cross-references: GB:M29872; NID:9178131
A:Note: the authors translated sequence upstream of Met-1
R:Hur, M.W.; Edenberg, H.J.
J. Biol. Chem. 270, 9002-9009, 1995
A:Title: Cell-specific function of cis-acting elements in the regulation of human alc

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:37:27 ; Search time 63.57 Seconds
(without alignments)
9.586 Million cell updates/sec

Title: 09-185908-1D
Perfect score: 35
Sequence: 1 WXXXXFXG 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	31	88.6	213	2	A56152	major 25k outer me
2	30	85.7	369	1	D64763	alcohol dehydrogen
3	30	85.7	369	1	S57525	alcohol dehydrogen
4	30	85.7	369	2	G85530	hypothetical prote
5	30	85.7	373	1	A33419	alcohol dehydrogen
6	30	85.7	373	1	DERPA	alcohol dehydrogen
7	30	85.7	373	1	S68061	alcohol dehydrogen
8	30	85.7	374	1	DEHRC2	alcohol dehydrogen
9	30	85.7	374	1	A56643	alcohol dehydrogen
10	30	85.7	376	1	JC4967	alcohol dehydrogen
11	30	85.7	376	1	S51187	alcohol dehydrogen
12	30	85.7	378	1	A49662	alcohol dehydrogen
13	30	85.7	378	1	H64052	alcohol dehydrogen
14	30	85.7	378	2	F81097	probable alcohol d
15	30	85.7	379	1	S51357	alcohol dehydrogen
16	30	85.7	379	1	S71244	alcohol dehydrogen
17	30	85.7	381	1	JN0447	alcohol dehydrogen
18	30	85.7	381	2	T03289	formaldehyde dehyd
19	30	85.7	381	2	T04164	formaldehyde dehyd
20	30	85.7	386	1	S31140	alcohol dehydrogen
21	30	85.7	396	2	S31959	acyl-lacetyl-carrier
22	30	85.7	733	1	S33643	transforming prote
23	29	82.9	199	1	TJBEKL	tail assembly prot
24	29	82.9	224	2	B85584	probable tail comp
25	29	82.9	367	2	T44888	probable aminometh
26	29	82.9	540	2	B83850	D-3-phosphoglycer
27	29	82.9	663	1	B54075	archidonate 12-11
28	29	82.9	663	1	S30051	archidonate 12-11
29	29	82.9	663	2	I52462	archidonate 12-11

30	29	82.9	1220	2	A56236	probable RNA helic
31	28	80.0	528	2	G70854	probable sera prot
32	28	80.0	528	2	T45418	phosphoglycerate d
33	27	77.1	159	2	S41178	gene 36 protein -
34	27	77.1	193	2	A37873	cerebellin precurs
35	27	77.1	224	2	A60032	cerebellin-like gl
36	27	77.1	280	2	A39484	androgen-withdraw
37	27	77.1	330	2	T34972	probable membrane
38	27	77.1	370	2	B83191	alcohol dehydrogen
39	27	77.1	599	2	A29947	prostaglandin-endo
40	27	77.1	600	2	A28960	prostaglandin-endo
41	27	77.1	600	2	S00561	probable sensory t
42	27	77.1	1639	2	T50119	hypothetical prote
43	26	74.3	93	2	H36808	hypothetical 11k p
44	26	74.3	100	2	J00859	unc1 protein homol
45	26	74.3	135	2	J01225	alcohol dehydrogen
46	26	74.3	185	2	S20593	hypothetical prote
47	26	74.3	198	2	G85506	probable tail comp
48	26	74.3	224	2	G85842	probable short cha
49	26	74.3	242	2	F83629	glycerol facilitat
50	26	74.3	264	2	H64086	

ALIGNMENTS

RESULT 1
A56152
major 25k outer membrane protein precursor - Brucella abortus
C:Species: Brucella abortus
C:Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 08-Oct-1999
C:Accession: A56152
R:de Wergifosse, P.; Lintermans, P.; Limet, J.N.; Claeckaert, A.
J. Bacteriol. 177, 1911-1914, 1995
A:Title: Cloning and nucleotide sequence of the gene coding for the major 25-kilodalton
A:Reference number: A56152; MUID:95204367
A:Accession: A56152
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-213 <DEA>
A:Cross-references: GB:X79284; NID:g769744; PIDN:CA55872.1; PID:g769745

Query Match 88.6%; Score 31; DB 2; Length 213;
Best Local Similarity 62.5%; Pred. No. 7.6;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 WXXXXFXG 8
DB 74 WKAGAFAG 81

RESULT 2
D64763
alcohol dehydrogenase (EC 1.1.1.1) C - Escherichia coli
N:Alternate names: class III alcohol dehydrogenase
N:Contains: formaldehyde dehydrogenase (glutathione) (EC 1.2.1.1)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 11-Jun-1999
C:Accession: D64763; A42015; S78608
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.: Rose, D.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: D64763
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-369 <BLAT>
A:Cross-references: GB:A0000142; GB:U00096; NID:q1786542; PIDN:AACT3459.1; PID:q17865
A:Experimental source: strain K-12, substrain MGL55
R:Guthell, W.G.; Holmquist, B.; Vallee, B.L.
Biochemistry 31, 475-481, 1992


```
; Sequence 5, Application US/08480065
; Patent No. 5837479
; GENERAL INFORMATION:
; APPLICANT: Young, Donald A.
; APPLICANT: O'Banion, Michael K.
; APPLICANT: Minn, Virginia D.
; TITLE OF INVENTION: SCREENIN ASSAYS FOR INHIBITORS OF
; TITLE OF INVENTION: MAMMALIAN PROSTAGLANDIN H SYNTHASE-2
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,065
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 3996-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 604 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-480-065-5

Query Match          74.3%; Score 26; DB 2; Length 604;
Best Local Similarity 50.0%; Pred. No. 3.6e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsfxxg 8
|| : ||
Db 531 WKPSTFGG 538

RESULT 20
US-08-487-744-2
; Sequence 2, Application US/08487744
; Patent No. 6048850
; GENERAL INFORMATION:
; APPLICANT: Young, Donald A.
; APPLICANT: O'Banion, Michael K.
; APPLICANT: Winn, Virginia D.
; TITLE OF INVENTION: METHODS FOR INHIBITING PROSTAGLANDIN
; TITLE OF INVENTION: SYNTHESIS IN A MAMMALIAN HOST
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,744
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 3996-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 604 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-487-744-2

Query Match          74.3%; Score 26; DB 3; Length 604;
Best Local Similarity 50.0%; Pred. No. 3.6e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsfxxg 8
|| : ||
Db 531 WKPSTFGG 538
```

Search completed: January 14, 2002, 07:23:43
Job time: 74 sec

ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,753
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 3996-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 604 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-487-753-5

Query Match 74.3% Score 26; DB 1; Length 604;
Best Local Similarity 50.0%; Pred. No. 3.6e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsfxg 8
||:| |
Db 531 WKPSFTFG 538

RESULT 17
US-08-480-065-2
Sequence 2, Application US/08480065
Patent No. 5837479
GENERAL INFORMATION:
APPLICANT: Young, Donald A.
APPLICANT: O'Banion, Michael K.
TITLE OF INVENTION: SCREENIN ASSAYS FOR INHIBITORS OF
TITLE OF INVENTION: MAMMALIAN PROSTAGLANDIN H SYNTHASE-2
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,065
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 3996-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 604 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-480-065-2

Query Match 74.3% Score 26; DB 2; Length 604;
Best Local Similarity 50.0%; Pred. No. 3.6e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsfxg 8
||:| |
Db 531 WKPSFTFG 538

RESULT 18
US-08-480-065-4
Sequence 4, Application US/08480065
Patent No. 5837479
GENERAL INFORMATION:
APPLICANT: Young, Donald A.
APPLICANT: O'Banion, Michael K.
TITLE OF INVENTION: SCREENIN ASSAYS FOR INHIBITORS OF
TITLE OF INVENTION: MAMMALIAN PROSTAGLANDIN H SYNTHASE-2
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,065
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 3996-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 604 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-480-065-4

Query Match 74.3% Score 26; DB 2; Length 604;
Best Local Similarity 50.0%; Pred. No. 3.6e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsfxg 8
||:| |
Db 531 WKPSFTFG 538

RESULT 19
US-08-480-065-5

TELECOMMUNICATION INFORMATION:
TELEPHONE: (908)594-3199
TELEFAX: (908)594-4720
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 604 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-064-271-10

Query Match 74.3%; Score 26; DB 1; Length 604;
Best Local Similarity 50.0%; Pred. No. 3.6e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxfxg 8
||:|
Db 531 WKPSTFGG 538

RESULT 14
US-08-487-753-2
; Sequence 2, Application US/08487753
; Patent No. 5807733
; GENERAL INFORMATION:
; APPLICANT: Young, Donald A.
; APPLICANT: O'Banion, Michael K.
; APPLICANT: Winn, Virginia D.
; TITLE OF INVENTION: MAMMALIAN PROSTAGLANDIN H SYNTHASE-2
; TITLE OF INVENTION: FUSION PROTEINS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/487,753
; APPLICATION NUMBER: US/08/487,753
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 3996-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 604 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-487-753-2

Query Match 74.3%; Score 26; DB 1; Length 604;
Best Local Similarity 50.0%; Pred. No. 3.6e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
OY 1 wxxxxfxg 8
||:|
Db 531 WKPSTFGG 538

RESULT 15
US-08-487-753-4
; Sequence 4, Application US/08487753
; Patent No. 5807733
; GENERAL INFORMATION:
; APPLICANT: Young, Donald A.
; APPLICANT: O'Banion, Michael K.
; APPLICANT: Winn, Virginia D.
; TITLE OF INVENTION: MAMMALIAN PROSTAGLANDIN H SYNTHASE-2
; TITLE OF INVENTION: FUSION PROTEINS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/487,753
; APPLICATION NUMBER: US/08/487,753
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 3996-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 604 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-487-753-4

Query Match 74.3%; Score 26; DB 1; Length 604;
Best Local Similarity 50.0%; Pred. No. 3.6e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxxfxg 8
||:|
Db 531 WKPSTFGG 538

RESULT 16
US-08-487-753-5
; Sequence 5, Application US/08487753
; Patent No. 5807733
; GENERAL INFORMATION:
; APPLICANT: Young, Donald A.
; APPLICANT: O'Banion, Michael K.
; APPLICANT: Winn, Virginia D.
; TITLE OF INVENTION: MAMMALIAN PROSTAGLANDIN H SYNTHASE-2
; TITLE OF INVENTION: FUSION PROTEINS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.

APPLICANT: Mituhuh, Vernon
TITLE OF INVENTION: MICROBIAL SYSTEM FOR FORMALDEHYDE
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/919,953
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/608,241
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J
REGISTRATION NUMBER: 27,386
REFERENCE/DOCKET NUMBER: 960296,93511
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 376 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-919-953-2

Query Match 77.1%; Score 27; DB 2; Length 376;
Best Local Similarity 50.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsfxx 8
DB 316 WKGSATFG 323

RESULT 11
US-09-192-983-2
Sequence 2, Application US/09192983A
GENERAL INFORMATION:
APPLICANT: Donohue, Timothy
APPLICANT: Barber, Robert
APPLICANT: Mituhuh, Vernon
TITLE OF INVENTION: Microbial System for Formaldehyde Sensing and
FILE REFERENCE: 960296,95505
CURRENT APPLICATION NUMBER: US/09/192,983A
CURRENT FILING DATE: 1998-11-16
EARLIER APPLICATION NUMBER: 08/919,953
EARLIER FILING DATE: 1997-08-29
EARLIER APPLICATION NUMBER: 08/608,241
EARLIER FILING DATE: 1996-02-28
NUMBER OF SEQ ID NOS: 7
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 2
LENGTH: 376
TYPE: PRT
ORGANISM: Rhodobacter sphaeroides
US-09-192-983-2

Query Match 77.1%; Score 27; DB 4; Length 376;
Best Local Similarity 50.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsfxx 8
DB 316 WKGSATFG 323

RESULT 12
US-09-214-278-1
Sequence 1, Application US/09214278
Patent No. 6291210
GENERAL INFORMATION:
APPLICANT: Sakano, Seiji
APPLICANT: Itoh, Akira
TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE
FILE REFERENCE: KP-8576
CURRENT APPLICATION NUMBER: US/09/214,278
CURRENT FILING DATE: 1999-01-26
NUMBER OF SEQ ID NOS: 32
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 1
LENGTH: 214
TYPE: PRT
ORGANISM: Homo sapiens
US-09-214-278-1

Query Match 74.3%; Score 26; DB 4; Length 214;
Best Local Similarity 50.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 wkxxsfxx 8
DB 152 WKSLEHFG 159

RESULT 13
US-08-064-271-10
Sequence 10, Application US/08064271
Patent No. 5543297
GENERAL INFORMATION:
APPLICANT: Kennedy, Brian P.
APPLICANT: Cromlish, Wanda A.
APPLICANT: Mancini, Joseph A.
APPLICANT: O'Neill, Gary
APPLICANT: Vickers, Philip J.
APPLICANT: Wong, Elizabeth
TITLE OF INVENTION: HUMAN CYCLOOXYGENASE-2 CDNA AND
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: 126 Lincoln Avenue
CITY: Rahway
STATE: NJ
COUNTRY: USA
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 in, 1.4kb
COMPUTER: Apple Macintosh
OPERATING SYSTEM: System 7
SOFTWARE: Microsoft Word 5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/064,271
FILING DATE: 19930506
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Panzer, Curtis C.
REGISTRATION NUMBER: 33,752
REFERENCE/DOCKET NUMBER: 189061A

```
RESULT 7
US-09-140-804-7
; Sequence 7, Application US/09140804
; Patent No. 6197930
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Humes, Jacqueline M.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-49
; CURRENT APPLICATION NUMBER: US/09/140,804
; CURRENT FILING DATE: 1998-08-26
; EARLIER APPLICATION NUMBER: 60/056,983
; EARLIER FILING DATE: 1997-08-26
; NUMBER OF SEQ. ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 7
; LENGTH: 222
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-140-804-7

Query Match          80.0%; Score 28; DB 4; Length 222;
Best Local Similarity 50.0%; Pred. No. 54;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxxxfxg 8
   ||:| |
Db 209 WKSYTFSG 216

RESULT 8
US-08-608-241-2
; Sequence 2, Application US/08608241
; Patent No. 5747328
; GENERAL INFORMATION:
; APPLICANT: Donohue, Timothy J
; APPLICANT: Barber, Robert D
; APPLICANT: Withuhn, Vernon
; TITLE OF INVENTION: MICROBIAL SYSTEM FOR FORMALDEHYDE
; TITLE OF INVENTION: SENSING AND REMEDIATION
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/608,241
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 960296.93511
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 376 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-608-241-2
```

```
Query Match          77.1%; Score 27; DB 1; Length 376;
Best Local Similarity 50.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxxxfxg 8
   ||:| |
Db 316 WKGSAFGG 323

RESULT 9
US-08-922-182-2
; Sequence 2, Application US/08922182
; Patent No. 5834300
; GENERAL INFORMATION:
; APPLICANT: Donohue, Timothy J
; APPLICANT: Barber, Robert D
; APPLICANT: Withuhn, Vernon
; TITLE OF INVENTION: MICROBIAL SYSTEM FOR FORMALDEHYDE
; TITLE OF INVENTION: SENSING AND REMEDIATION
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/922,182
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/608,241
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 960296.93511
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 376 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-922-182-2

Query Match          77.1%; Score 27; DB 2; Length 376;
Best Local Similarity 50.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxxxfxg 8
   ||:| |
Db 316 WKGSAFGG 323

RESULT 10
US-08-919-953-2
; Sequence 2, Application US/08919953
; Patent No. 5837481
; GENERAL INFORMATION:
; APPLICANT: Donohue, Timothy J
; APPLICANT: Barber, Robert D
```

EARLIER FILING DATE: 1998-12-11
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 345
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-222-786-2

Query Match 85.7%: Score 30; DB 4; Length 345;
Best Local Similarity 62.5%: Pred. No. 31;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsfxg 8
DB 132 WKRSSFNG 139

RESULT 3
US-09-222-817-12
Sequence 12, Application US/09222817
Patent No. 6037154
GENERAL INFORMATION:
APPLICANT: MIKIKO SUGA, Masakazu SUGIMOTO, Tsuyoshi OSUMI, Tsuyoshi NAKAMATSU,
FILE REFERENCE: OP813
TITLE OF INVENTION: METHOD OF PRODUCING L-SERINE BY FERMENTATION
CURRENT APPLICATION NUMBER: US/09/222, 817
EARLIER FILING DATE: 1998-12-30
EARLIER APPLICATION NUMBER: JP 10-3751
EARLIER FILING DATE: 1998-01-12
EARLIER APPLICATION NUMBER: JP 10-353521
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 12
LENGTH: 530
TYPE: PRT
ORGANISM: Brevibacterium flavum
US-09-222-817-12

Query Match 85.7%: Score 30; DB 3; Length 530;
Best Local Similarity 62.5%: Pred. No. 46;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsfxg 8
DB 132 WKRSSFNG 139

RESULT 4
US-09-222-817-14
Sequence 14, Application US/09222817
Patent No. 6037154
GENERAL INFORMATION:
APPLICANT: MIKIKO SUGA, Masakazu SUGIMOTO, Tsuyoshi OSUMI, Tsuyoshi NAKAMATSU,
FILE REFERENCE: OP813
TITLE OF INVENTION: METHOD OF PRODUCING L-SERINE BY FERMENTATION
CURRENT APPLICATION NUMBER: US/09/222, 817
EARLIER FILING DATE: 1998-12-30
EARLIER APPLICATION NUMBER: JP 10-3751
EARLIER FILING DATE: 1998-01-12
EARLIER APPLICATION NUMBER: JP 10-353521
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 14
LENGTH: 530
TYPE: PRT
ORGANISM: Brevibacterium flavum
US-09-222-817-14

Query Match 85.7%: Score 30; DB 3; Length 530;
Best Local Similarity 62.5%: Pred. No. 46;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsfxg 8
DB 132 WKRSSFNG 139

RESULT 5
US-09-222-786-12
Sequence 12, Application US/09222786A
Patent No. 6258573
GENERAL INFORMATION:
APPLICANT: MIKIKO SUGA, Masakazu SUGIMOTO, Tsuyoshi OSUMI, Tsuyoshi NAKAMATSU,
FILE REFERENCE: OP812
TITLE OF INVENTION: METHOD OF PRODUCING L-SERINE BY FERMENTATION
CURRENT APPLICATION NUMBER: US/09/222, 786A
EARLIER FILING DATE: 1998-12-30
EARLIER APPLICATION NUMBER: JP 10-3751
EARLIER FILING DATE: 1998-01-12
EARLIER APPLICATION NUMBER: JP 10-353513
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 12
LENGTH: 530
TYPE: PRT
ORGANISM: Brevibacterium flavum
US-09-222-786-12

Query Match 85.7%: Score 30; DB 4; Length 530;
Best Local Similarity 62.5%: Pred. No. 46;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsfxg 8
DB 132 WKRSSFNG 139

RESULT 6
US-09-222-786-14
Sequence 14, Application US/09222786A
Patent No. 6258573
GENERAL INFORMATION:
APPLICANT: MIKIKO SUGA, Masakazu SUGIMOTO, Tsuyoshi OSUMI, Tsuyoshi NAKAMATSU,
FILE REFERENCE: OP812
TITLE OF INVENTION: METHOD OF PRODUCING L-SERINE BY FERMENTATION
CURRENT APPLICATION NUMBER: US/09/222, 786A
EARLIER FILING DATE: 1998-12-30
EARLIER APPLICATION NUMBER: JP 10-3751
EARLIER FILING DATE: 1998-01-12
EARLIER APPLICATION NUMBER: JP 10-353513
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 14
LENGTH: 530
TYPE: PRT
ORGANISM: Brevibacterium flavum
US-09-222-786-14

Query Match 85.7%: Score 30; DB 4; Length 530;
Best Local Similarity 62.5%: Pred. No. 46;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsfxg 8
DB 132 WKRSSFNG 139

Gencore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:23:42 ; Search time 48.05 Seconds
(without alignments)
3.747 Million cell updates/sec

Title: 09-185908-1c
Perfect score: 35
Sequence: 1 wkxxsfxxg 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues
Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCrUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfilest.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	30	85.7	345	3	US-09-222-817-2
2	30	85.7	345	4	US-09-222-786-2
3	30	85.7	530	3	US-09-222-817-12
4	30	85.7	530	3	US-09-222-817-14
5	30	85.7	530	4	US-09-222-786-12
6	30	85.7	530	4	US-09-222-786-14
7	28	80.0	222	4	US-09-140-804-7
8	27	77.1	376	1	US-08-608-241-2
9	27	77.1	376	2	US-08-922-182-2
10	27	77.1	376	2	US-08-919-953-2
11	27	77.1	376	4	US-09-192-983-2
12	26	74.3	214	4	US-09-214-278-1
13	26	74.3	604	1	US-08-064-271-10
14	26	74.3	604	1	US-08-487-753-2
15	26	74.3	604	1	US-08-487-753-4
16	26	74.3	604	1	US-08-487-753-5
17	26	74.3	604	1	US-08-480-065-2
18	26	74.3	604	2	US-08-480-065-4
19	26	74.3	604	2	US-08-480-065-5
20	26	74.3	604	3	US-08-487-744-2
21	26	74.3	604	3	US-08-487-744-4
22	26	74.3	604	3	US-08-487-744-5
23	26	74.3	604	3	US-08-930-589A-18
24	26	74.3	604	3	PCT-US93-09167-2
25	26	74.3	604	5	PCT-US93-09167-4
26	26	74.3	604	5	PCT-US93-09167-5
27	26	74.3	759	2	US-08-637-759B-89

28	26	74.3	759	3	US-08-871-355A-89	Sequence 89, Appl
29	26	74.3	1043	2	US-08-724-354D-4	Sequence 4, Appl
30	26	74.3	1043	3	US-09-270-984A-4	Sequence 4, Appl
31	26	74.3	1055	4	US-09-214-278-2	Sequence 2, Appl
32	26	74.3	1065	2	US-08-400-159-8	Sequence 8, Appl
33	26	74.3	1118	2	US-08-724-354D-2	Sequence 2, Appl
34	26	74.3	1118	3	US-09-270-984A-2	Sequence 2, Appl
35	26	74.3	1148	4	US-08-882-046-4	Sequence 4, Appl
36	26	74.3	1212	4	US-09-214-278-3	Sequence 3, Appl
37	26	74.3	1238	4	US-09-214-278-5	Sequence 5, Appl
38	26	74.3	1248	4	US-08-882-046-6	Sequence 6, Appl
39	26	74.3	1257	3	US-08-611-729A-8	Sequence 8, Appl
40	25	71.4	24	3	US-08-504-538A-8	Sequence 8, Appl
41	25	71.4	24	4	US-09-249-458A-8	Sequence 8, Appl
42	25	71.4	24	5	PCT-US95-09307-8	Sequence 8, Appl
43	25	71.4	34	2	US-08-751-767A-38	Sequence 38, Appl
44	24	68.6	99	4	US-08-985-050-4	Sequence 4, Appl
45	24	68.6	123	4	US-09-357-251-22	Sequence 22, Appl
46	24	68.6	154	4	US-09-193-104-16	Sequence 16, Appl
47	24	68.6	218	2	US-08-399-889-25	Sequence 25, Appl
48	24	68.6	218	3	US-09-167-364-25	Sequence 4, Appl
49	24	68.6	218	4	US-09-439-897-4	Sequence 2, Appl
50	24	68.6	228	1	US-08-447-591-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-222-817-2
; Sequence 2, Application US/09222817
; Patent No. 6037154
; GENERAL INFORMATION:
; APPLICANT: Mikiko SUGA, Masakazu SUGIMOTO, Tsuyoshi OSUMI, Tsuyoshi NAKAMATSU,
; TITLE OF INVENTION: METHOD OF PRODUCING L-SERINE BY FERMENTATION
; FILE REFERENCE: OP813
; CURRENT FILING DATE: 1998-12-30
; EARLIER APPLICATION NUMBER: JP 10-3751
; EARLIER FILING DATE: 1998-01-12
; EARLIER APPLICATION NUMBER: JP 10-353521
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; US-09-222-817-2

Query Match 85.7%; Score 30; DB 3; Length 345;
Best local Similarity 62.5%; Pred. NO. 31;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsfxxg 8
DB 132 WKRSSFNG 139

RESULT 2
US-09-222-786-2
; Sequence 2, Application US/09222786A
; Patent No. 6258573
; GENERAL INFORMATION:
; APPLICANT: Mikiko SUGA, Masakazu SUGIMOTO, Tsuyoshi OSUMI, Tsuyoshi NAKAMATSU,
; TITLE OF INVENTION: METHOD OF PRODUCING L-SERINE BY FERMENTATION
; FILE REFERENCE: OP812
; CURRENT APPLICATION NUMBER: US/09/222,786A
; CURRENT FILING DATE: 1998-12-30
; EARLIER APPLICATION NUMBER: JP 10-3751
; EARLIER FILING DATE: 1998-01-12
; EARLIER APPLICATION NUMBER: JP 10-353513

CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.
XX
SQ Sequence 126 AA;

Query Match 80.0%; Score 28; DB 22; Length 126;
Best Local Similarity 50.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsfxxg 8
|| | : |
Db 73 wkqssyag 80

RESULT 20
AAM42193
ID AAM42193 standard; Protein; 126 AA.
XX
AC AAM42193;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 7124.
XX
KM Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;
KM peripheral nervous system; neuropathy; central nervous system; CNS;
KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KM chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KM leukaemia.
XX
CS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
XX
DR N-PSDB; AAI61349.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 2; SEQ ID NO 7124; 10078bp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM4213) with noctropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 126 AA;

Query Match 80.0%; Score 28; DB 22; Length 126;
Best Local Similarity 50.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsfxxg 8
|| | : |
Db 73 wkqssyag 80

Search completed: January 14, 2002, 07:36:00
Job time: 811 sec

Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 wxxxxfxg 8
 11 | : |
 Db 3 wktssyvg 10

RESULT 18
 AAY12226
 ID AAY12226 standard; Protein: 114 AA.

AC AAY12226;
 DT 18-JUN-1999 (first entry)

Human 5' EST secreted protein SEQ ID NO: 539.

Human; secreted protein; EST; expressed sequence tag; diagnosis;
 forensic; gene therapy; chromosome mapping; signal peptide;
 upstream regulatory sequence; cytokine activity; cell proliferation;
 differentiation; haematopoiesis regulation; tissue growth regulation;
 reproductive hormone regulation; chemotactic; chemokine; haemostatic;
 thrombolytic; anti-inflammatory; tumour inhibition.

Homo sapiens.

MO9906554-A2.

11-FEB-1999.

31-JUL-1998; 98WO-1B01238.

01-AUG-1997; 97US-0905134.

(GEST) GENSET.

Duclert A, Dumas Malne Edwards J, Lacroix B;

WPI: 1999-153784/13.

N-PSDB: AAX41059.

New nucleic acids encoding human secreted proteins - obtained from
 cDNA libraries prepared from kidney, fetal kidney, dystrophic
 muscle, muscle and heart tissue

Claim 34; Page 600-601; 622pp; English.

AAV40826 to AAX41093 represent 5' expressed sequence tags (ESTs) for
 human secreted proteins, and encode the proteins given in AAY01602 and
 AAY11994 to AAY12260, respectively. The proteins given represent the
 signal peptide and an N-terminal fragment of a secreted protein. The
 nucleic acid sequences can be used for producing secreted human gene
 products. They can also be used to develop products for diagnosis and
 therapy. The proteins obtained may have cytokine activity, cell
 proliferation/differentiation activity, haematopoiesis regulating
 activity, tissue growth regulating activity, reproductive hormone
 regulating activity, chemotactic/chemokine activity, haemostatic and
 thrombolytic activity, receptor/ligand activity, anti-inflammatory
 activity, tumour inhibition activity or other activities. The products
 can be used in forensic, gene therapy and chromosome mapping procedures.
 The sequences can also be used for obtaining corresponding promoter
 sequences. The nucleic acids encoding the signal peptide can be used
 for directing extracellular secretion of a polypeptide or the insertion
 of a polypeptide into a membrane, or importing a polypeptide into
 a cell.

Sequence 114 AA;

Query Match 80.0%; Score 28; DB 20; Length 114;
 Best Local Similarity 50.0%; Pred. No. 96;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxxxfxg 8
 11 | : |
 Db 30 wktssyvg 37

RESULT 19
 AAM25829
 ID AAM25829 standard; Protein: 126 AA.

AAM25829;

16-OCT-2001 (first entry)

Human protein sequence SEQ ID NO:1344.

Human; cancer; HIV infection; human immunodeficiency virus;
 antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
 antibacterial; endocrine; cardiant; central nervous system; virucide;
 anti-HIV; fungicide; antitumagen; cardiovascular; antianemic; anaemia;
 antiagregant; haemostatic; vulnary; antitumor; osteopathic; eczema;
 dermatological; antiallergic; antidiabetic; antidiabetic; cytostatic;
 neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
 immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
 immunosuppressive; rheumatoid arthritis; septic shock; pancreatitis;
 antianaphylactic; neuropathology; cardiac anaphylaxis; autoimmunity;
 cardiac dysfunction; neuropathology; cardiac anaphylaxis; asthma;
 genetic disease; haematopoietic disorder; platelet disorder; asthma;
 thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
 allergic rhinitis; diabetes; multiple sclerosis; depression;
 Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 neurological disorder.

Homo sapiens.

WO200153455-A2.

26-JUL-2001.

22-DEC-2000; 2000WO-US35017.

23-DEC-1999; 99US-0471275.

21-JAN-2000; 2000US-0488725.

25-APR-2000; 2000US-0552317.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT;

WPI: 2001-457603/49.

N-PSDB: AAH99770.

Isolated human polynucleotides encoding polypeptides, useful for the

treatment and diagnosis of e.g. cancer, ulcers and HIV infection -

Claim 20; Page 278; 1217pp; English.

AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
 AAM25963. The proteins can have activities based on the tissues and
 cells they are expressed in, such as: antiinflammatory; antirheumatic;
 antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
 central nervous system; virucide; anti-HIV; fungicide; antitumagen;
 cardiovascular; antianemic; antiagregant; haemostatic; vulnary;
 antitumor; osteopathic; dermatological; antiallergic; antidiabetic;
 antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
 antiparkinsonian; and immunostimulant. The proteins and polynucleotides
 encoding them can be used in gene therapy, antisense therapy and vaccine
 production. The proteins and polynucleotides are useful for screening for
 agonists or antagonists of a protein and for the treatment and diagnosis
 of disorders associated with the activity of a protein e.g. inflammation,
 rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
 neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
 infections, autoimmunity, genetic diseases, haematopoietic disorders,
 anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,

KM cell culture; chemotaxis; food additive;
KW binding partner identification.
XX
XX Homo sapiens.
XX MO200136432-A2.
XX
XX 25-MAY-2001.
XX
XX 15-NOV-2000; 2000WO-US31162.
XX PF 19-NOV-1999; 990US-0166415.
XX PR 30-JUN-2000; 2000US-0215136.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Komatsoulis GA, Baker KP, Young PE;
XX WPI; 2001-343793/36.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is
PT used in preventing, treating or ameliorating a medical condition -
XX
XX Disclosure; Page 39; 509pp; English.
XX
XX AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted
CC protein genes, and AAE04199-AAE04239 represent the proteins they encode.
CC AAE04240-AAE04297 represent human secreted protein fragments or variants.
CC The secreted proteins and their genes are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 18 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders, schizophrenia, asthma,
CC Parkinson's disease, cognitive disorders, schizophrenia, atherosclerosis,
CC skin disorders (e.g., psoriasis), sepsis, diabetes, kidney disorders,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin ageing due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein fragment referred to in the disclosure of the invention.
XX
XX Sequence 56 AA:
SO

Query Match 80.0%; Score 28; DB 22; Length 56;
Best Local Similarity 50.0%; Pred. No. 49;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsfxg 8
|| : |
DB 6 wksysyvg 13

RESULT 15
AAAY32938
ID AAY32938 standard; Protein: 72 AA.
XX
XX AAY32938;
XX

DT 04-NOV-1999 (first entry)
XX
XX Human cerebellin-2 protein sequence.
DE
XX
XX Cerebellin-2; human; neurological disorder; synapse function disruption;
KW Parkinson's disease; Alzheimer's disease; affective disorder; therapy;
KW bipolar disorder; unipolar disorder; schizophrenia; Shy-Drager syndrome;
KW olivopontocerebellar atrophy; vaccine production; antibody production.
XX
XX Homo sapiens.
XX OS
XX
XX WO9942576-A1.
XX PN
XX 26-AUG-1999.
XX PD
XX
XX 22-FEB-1999; 99WO-GB00534.
XX PE
XX 22-JAN-1999; 99GB-0001463.
XX PR 23-FEB-1998; 98GB-0003786.
XX
XX (SMIK) SMITHKLINE BEECHAM PLC.
XX
XX Barnes MR;
XX
XX WPI; 1999-527473/44.
XX DR N-PSDB; AAZ11193.
XX
XX Isolated cerebellin-2 related polypeptides useful for diagnosing and
PT treating neurological disorders such as Parkinson's disease,
PT Alzheimer's disease and schizophrenia
XX
XX Claim 14; Page 30; 35pp; English.
XX
XX This sequence represents the human cerebellin-2 of the invention. The
XX cerebellin-2 protein may be administered to treat or prevent neurological
XX disorders associated with the inappropriate expression of cerebellin-2
XX proteins and disruption of the synapse function. These disorders include
XX Parkinson's disease, Alzheimer's disease, affective disorders (both
XX bipolar and unipolar disorders), schizophrenia, olivopontocerebellar
XX atrophy, Shy-Drager syndrome and other disorders caused by disruption of
XX synapse function. Cerebellin-2 may also be used as an antigen in the
XX production of vaccines and antibodies specific for the protein and in
XX assays to identify agonists and other antagonists of cerebellin-2
XX function.
XX
XX Sequence 72 AA:
SO

Query Match 80.0%; Score 28; DB 20; Length 72;
Best Local Similarity 50.0%; Pred. No. 62;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsfxg 8
|| : |
DB 59 wksylsfsg 66

RESULT 16
AAM40407
ID AAM40407 standard; Protein: 83 AA.
XX
XX AAM40407;
XX
XX 22-OCT-2001 (first entry)
XX DE Human polypeptide SEQ ID NO 3552.
XX
XX Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.

KW reproductive disorder; gastrointestinal disorder; renal disorder;
 KW infectious disease; cardiovascular disorder.
 XX
 OS Homo sapiens.
 XX
 PN W0200055351-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05883.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI: 2000-587534/55.
 DR N-PSDB; AAC98643.
 XX
 PT Colon cancer associated gene sequences, referred to as colon cancer
 PT antigens, useful for the treatment, prevention, and diagnosis of colon
 PT disorders such as colon cancer -
 XX
 PS Claim 11; Page 1990; 2104pp; English.
 XX
 CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
 CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
 CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
 CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
 CC vulnary, nephrotoxic, antineoplastic and antibacterial activities, and
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,
 CC proteins and antibodies to the proteins are useful for the prevention,
 CC treatment and diagnosis of colon disorders, such as colon cancer. The
 CC polynucleotides may be used in diagnostics and research, such as for
 CC chromosome identification, and as hybridisation probes. The proteins
 CC may also be used to prevent diseases such as neural disorders, immune
 CC system disorders, muscular disorders, reproductive disorders,
 CC gastrointestinal disorders, wounds, renal disorders, infectious
 CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
 CC AAB54007 represent sequences used in the exemplification of the present
 CC invention.
 CC
 SO Sequence 75 AA:

Query Match 82.9%; Score 29; DB 21; Length 75;
 Best Local Similarity 75.0%; Pred. No. 40;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 wxxsfxg 8
 II III I
 Db 60 wlxsflyg 67

RESULT 13
 AAY12227
 ID AAY12227 standard; Protein: 55 AA.
 XX
 AC AAY12227;
 XX
 DT 18-JUN-1999 (first entry)
 XX
 DE Human 5' EST secreted protein SEQ ID NO: 540.
 XX
 KW Human: secreted protein; EST; expressed sequence tag; diagnosis;
 KW forensic; gene therapy; chromosome mapping; signal peptide;
 KW upstream regulatory sequence; cytokine activity; cell proliferation;
 KW differentiation; hematopoiesis regulation; tissue growth regulation;
 KW reproductive hormone regulation; chemotactic; chemokine; haemostatic;
 KW thrombolytic; anti-inflammatory; tumour inhibition.
 KW
 XX

OS Homo sapiens.
 XX
 PN W0906554-A2.
 XX
 PD 11-FEB-1999.
 XX
 PF 31-JUL-1998; 98MO-IB01238.
 XX
 PR 01-AUG-1997; 97US-0905134.
 XX
 PA (GSEST) GENSET.
 XX
 PI Duclert A, Dumas Milne Edwards J, Lacroix B;
 XX
 DR WPI: 1999-153784/13.
 DR N-PSDB; AAX41060.
 XX
 PT New nucleic acids encoding human secreted proteins - obtained from
 PT cDNA libraries prepared from kidney, fetal kidney, dystrophic
 PT muscle, muscle and heart tissue
 XX
 PS Claim 34; Page 601; 622pp; English.
 XX
 CC AAX40826 to AAX41093 represent 5' expressed sequence tags (ESTs) for
 CC human secreted proteins, and encode the proteins given in AAY01602 and
 CC AAY11994 to AAY12260, respectively. The proteins given represent the
 CC signal peptide and an N-terminal fragment of a secreted protein. The
 CC nucleic acid sequences can be used for producing secreted human gene
 CC products. They can also be used to develop products for diagnosis and
 CC therapy. The proteins obtained may have cytokine activity, cell
 CC proliferation/differentiation activity, hematopoiesis regulating
 CC activity, tissue growth regulating activity, reproductive hormone
 CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
 CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, tumour inhibition activity or other activities. The products
 CC can be used in forensic, gene therapy and chromosome mapping procedures.
 CC The sequences can also be used for obtaining corresponding promoter
 CC sequences. The nucleic acids encoding the signal peptide can be used
 CC for directing extracellular secretion of a polypeptide or the insertion
 CC of a polypeptide into a membrane, or importing a polypeptide into
 CC a cell.
 CC
 SO Sequence 55 AA:

Query Match 80.0%; Score 28; DB 20; Length 55;
 Best Local Similarity 50.0%; Pred. No. 48;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxsfxg 8
 II I: I
 Db 30 wktssyvg 37

RESULT 14
 AAE04286
 ID AAE04286 standard; peptide; 56 AA.
 XX
 AC AAE04286;
 XX
 DT 09-AUG-2001 (first entry)
 XX
 DE Human gene 10 encoded secreted protein fragment, SEQ ID NO:150.
 XX
 KW Human: secreted protein; proliferative disorder; cancer; tumour;
 KW foetal abnormality; developmental abnormality; hematopoietic disorder;
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;
 KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;
 KW endocrine disorder; infection; wound healing; vulnary;
 KW

Query Match 85.7%; Score 30; DB 20; Length 530;
Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 wkxxsfxx 8
DB 132 wkrsfng 139

RESULT 10
AA22647
ID AAY22647 standard; Protein; 530 AA.
XX
AC AAY22647;
XX
DT 02-SEP-1999 (first entry)
XX
DE Mutant 3-PGDH protein, also known as serA.
XX
KM Coryneform bacterium: L-serine production; phosphoserine phosphatase;
XX phosphoserine transaminase; large-scale fermentation; 3-PGDH; serA.
XX Brevibacterium flavum.
XX OS EP931833-A2.
XX PN 28-JUL-1999.
XX PD 12-JAN-1999; 99EP-0100325.
XX PF 11-DEC-1998; 98JP-0353521.
XX PR 12-JAN-1998; 98JP-0003751.
XX
XX (AJIN) AJINOMOTO CO INC.
XX
XX Hibino W, Ito M, Nakamatsu T, Osumi T, Suga M;
XX Sugimoto M;
XX WPI: 1999-397161/34.
XX DR N-PSDB: AAX81849.
XX
XX New coryneform bacterial strain useful for producing L-serine
XX
XX Example 5; Page 25-27; 33pp: English.
XX
XX The specification describes a coryneform bacterium which is able to
XX produce L-serine. The activity of at least one phosphoserine phosphatase
XX and phosphoserine transaminase is enhanced in the organism. The
XX organism is used for large-scale fermentation of L-serine for
XX amino acid mixtures which are used in pharmaceuticals, chemicals and
XX cosmetics. The present sequence represents a mutant 3-PGDH protein, also
XX known as serA, and is used in the course of the invention.
XX
XX Sequence 530 AA:
SQ

Query Match 85.7%; Score 30; DB 20; Length 530;
Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 wkxxsfxx 8
DB 132 wkrsfng 139

RESULT 11
AAG91161
ID AAG91161 standard; Protein; 530 AA.
XX
AC AAG91161;
XX
DT 26-SEP-2001 (first entry)
XX

DE C glutamicum protein fragment SEQ ID NO: 4915.
XX
XX Coryneform bacterium: amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis.
XX
XX Corynebacterium glutamicum.
XX OS EP1108790-A2.
XX PN 20-JUN-2001.
XX PD 18-DEC-2000; 2000EP-0127688.
XX PF 16-DEC-1999; 99JP-0377484.
XX PR 07-APR-2000; 2000JP-0159162.
XX PR 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX WPI: 2001-376931/40.
XX DR N-PSDB: AAH6380.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
XX mutation point of a gene, measuring expression of a gene, analysing
XX expression profile or pattern of a gene and identifying homologous gene
XX
XX Claim 17; SEQ ID NO: 4915; 246pp + Sequence Listing; English.
XX
XX

The present invention provides a number of nucleotide and protein
sequences from the Coryneform bacterium Corynebacterium glutamicum. These
are useful for identifying the mutation point of a gene derived from a
mutant of coryneform bacterium, measuring expression amount and
analysing the expression profile or expression pattern of a gene derived
from Coryneform bacterium, and identifying a homologue of a gene derived
from Coryneform bacterium. Coryneform bacteria are useful for producing
amino acids, nucleic acids, vitamins, saccharides and organic acids,
particularly L-lysine. The present sequence is a protein described
in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from the
European Patent Office.
CC
XX

Sequence 530 AA:
SQ

Query Match 85.7%; Score 30; DB 22; Length 530;
Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 wkxxsfxx 8
DB 132 wkrsfng 139

RESULT 12
AAB53886
ID AAB53886 standard; Protein; 75 AA.
XX
AC AAB53886;
XX
DT 09-MAR-2001 (first entry)
XX
XX Human colon cancer antigen protein sequence SEQ ID NO:1426.
XX
XX Human: colon cancer; colon cancer antigen; diagnosis; detection;
XX identification; cytostatic; cardioprotective; neuroprotective; vulnerrary;
XX immunomodulatory; muscular; gynaecological; gastrointestinal;
XX nephrotropic; antiinfective; antibacterial; gene therapy; wound;
XX neural disorder; immune system disorder; muscular disorder;
XX

CC The specification describes a coryneform bacterium which is able to
CC produce L-serine. The activity of at least one phosphoserine phosphatase
CC and phosphoserine transaminase is enhanced in the organism. The
CC organism is used for large-scale fermentation of L-serine for
CC amino acid mixtures which are used in pharmaceuticals, chemicals and
CC cosmetics. The present sequence represents 3-PGDH protein, also known
CC as serA, and is used in the course of the invention.

Sequence 530 AA;

```

RESULT      5
AA31651
ID  AAY31651 standard; Protein: 345 AA.
XX
XX  AAY31651;
AC
XX  09-NOV-1999 (first entry)
DT
XX  Corynebacterium glutamicum D-3-phosphoglycerate dehydrogenase.
DE
XX  D-3-phosphoglycerate dehydrogenase; sera gene; 3-PGDH; L-serine.
XX
XX  Corynebacterium glutamicum.
OS
XX  EP943687-A2.
PN
XX  22-SEP-1999.
PD
XX  12-JAN-1999; 99EP-0100324.
PF
XX  11-DEC-1998; 98JP-0353513.
PR  12-JAN-1998; 98JP-0003751.
XX
XX  (AJIN ) AJINOMOTO CO INC.
PA
XX  Hibino W, Ito M, Nakamatsu T, Osumi T, Suga M;
PI  Sugimoto M;
XX
XX  WPI; 1999-510578/43.
DR  N-PSDB; AAX87798.
XX
XX  New coryneform bacterium, useful for production of L-serine for
PT  pharmaceuticals, chemicals and cosmetics
XX
XX  Example 5; Page 13-14; 27pp; English.
XX
XX  This sequence represents the N-terminal portion of the
CC  D-3-phosphoglycerate dehydrogenase (3-PGDH) of Corynebacterium
CC  glutamicum strain K82 (FERM BP-2444), as deduced from the
CC  nucleotide sequence of PCR-amplified K82 DNA (see AAX87798). PCR
CC  primers (see AAX87799 and AAX87800) based on the amplified K82 DNA were
CC  used in the amplification of Brevibacterium flavum 3-PGDH sera
CC  gene sequences. The invention provides a method of producing
CC  L-serine, potentially on an industrial scale, using a coryneform
CC  bacterium having resistance to azaserine or beta-(2-thienyl)-
CC  DL-alanine and L-serine productivity. Such bacteria produce a
CC  3-PGDH (see AAY31650) in which feedback inhibition by L-serine is
CC  desensitised.
XX
XX  Sequence 345 AA;
SQ

Query Match      85.7%; Score 30; DB 20; Length 345;
Best Local Similarity 62.5%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY  1 wkxxsfng 8
DB  132 wkrsfng 139

RESULT      6
AAY22645
ID  AAY22645 standard; Protein: 345 AA.
XX
XX  AAY22645;
AC
XX  02-SEP-1999 (first entry)
DT
XX  3-PGDH protein, also known as sera.
DE
XX  Coryneform bacterium; L-serine production; phosphoserine phosphatase;
XX  phosphoserine transaminase; large-scale fermentation; 3-PGDH; sera.
KW

```

```

XX  Corynebacterium glutamicum.
OS
XX  EP931833-A2.
PN
XX  28-JUL-1999.
PD
XX  12-JAN-1999; 99EP-0100325.
PF
XX  11-DEC-1998; 98JP-0353521.
PR  12-JAN-1998; 98JP-0003751.
XX
XX  (AJIN ) AJINOMOTO CO INC.
PA
XX  Hibino W, Ito M, Nakamatsu T, Osumi T, Suga M;
PI  Sugimoto M;
XX
XX  WPI; 1999-397161/34.
DR  N-PSDB; AAX81839.
XX
XX  New coryneform bacterial strain useful for producing L-serine
PT  Example 5; Page 15-16; 33pp; English.
XX
XX  The specification describes a coryneform bacterium which is able to
CC  produce L-serine. The activity of at least one phosphoserine phosphatase
CC  and phosphoserine transaminase is enhanced in the organism. The
CC  organism is used for large-scale fermentation of L-serine for
CC  amino acid mixtures which are used in pharmaceuticals, chemicals and
CC  cosmetics. The present sequence represents 3-PGDH protein, also known
CC  as sera, and is used in the course of the invention.
XX
XX  Sequence 345 AA;
SQ

```

```

Query Match      85.7%; Score 30; DB 20; Length 345;
Best Local Similarity 62.5%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY  1 wkxxsfng 8
DB  132 wkrsfng 139

RESULT      7
AAY31650
ID  AAY31650 standard; Protein: 530 AA.
XX
XX  AAY31650;
AC
XX  09-NOV-1999 (first entry)
DT
XX  Brevibacterium flavum mutant D-3-phosphoglycerate dehydrogenase.
DE
XX  D-3-phosphoglycerate dehydrogenase; sera gene; 3-PGDH; L-serine;
XX  mutant.
XX  Brevibacterium flavum.
OS  Synthetic.
OS
XX  EP943687-A2.
PN
XX  22-SEP-1999.
PD
XX  12-JAN-1999; 99EP-0100324.
PF
XX  11-DEC-1998; 98JP-0353513.
PR  12-JAN-1998; 98JP-0003751.
XX
XX  (AJIN ) AJINOMOTO CO INC.
PA
XX  Hibino W, Ito M, Nakamatsu T, Osumi T, Suga M;
PI  Sugimoto M;

```

PR 06-JUN-1997; 97US-0048876.
 PR 06-JUN-1997; 97US-0048880.
 PR 06-JUN-1997; 97US-0048883.
 PR 06-JUN-1997; 97US-0048892.
 PR 06-JUN-1997; 97US-0048895.
 PR 06-JUN-1997; 97US-0048898.
 PR 06-JUN-1997; 97US-0048901.
 PR 06-JUN-1997; 97US-0048917.
 PR 06-JUN-1997; 97US-0048963.
 PR 06-JUN-1997; 97US-0048973.
 PR 06-JUN-1997; 97US-0049019.
 PR 06-JUN-1997; 97US-0049374.
 PR 05-SEP-1997; 97US-0057627.
 PR 05-SEP-1997; 97US-0057634.
 PR 05-SEP-1997; 97US-0057643.
 PR 05-SEP-1997; 97US-0057646.
 PR 05-SEP-1997; 97US-0057649.
 PR 05-SEP-1997; 97US-0057653.
 PR 05-SEP-1997; 97US-0057656.
 PR 05-SEP-1997; 97US-0057760.
 PR 05-SEP-1997; 97US-0057763.
 PR 05-SEP-1997; 97US-0057769.
 PR 05-SEP-1997; 97US-0057774.
 PR 05-SEP-1997; 97US-0057777.
 (HUMA-) HUMAN GENOME SCI INC.

PA Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA;
 PI Fan P, Feng P, Ferric AM, Fischer CL, Florence C;
 PI Florence K, Greene JM, Hu J, Kyaw H, Lallieur DW;
 PI Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM;
 PI Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;
 WI: 1999-059865/05.
 DR N-PSDB; AAV84624.

PT New isolated human genes and the secreted polypeptides they encode -
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 PS Claim 11; Page 589-590; 772pp; English.

XX The invention relates to nucleic acid sequences (AAV8411 to AAV84633)
 CC encoding human secreted proteins (AAW88534 to AAW88756). The secreted
 CC protein gene sequences are deposited with the ATCC under deposit numbers
 CC ATCC 97979, 97974, 97975, 97976, 97977, 209007, 209008, 209009, 209010,
 CC 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host
 CC cells comprising recombinant vectors containing the nucleic acid
 CC sequences are used for the recombinant production of the secreted
 CC proteins. The polynucleotide and amino acid sequences are useful for are
 CC useful for preventing, treating or ameliorating medical conditions e.g.
 CC by protein or gene therapy. Pathological conditions can be also
 CC diagnosed by determining the amount of the new polypeptides in a sample
 CC or by determining the presence of mutations in the new polynucleotides.
 CC Specific uses are described for each of the polynucleotides, based on
 CC which tissues they are most highly expressed in, and include developing
 CC products for the diagnosis or treatment of cancer, neurodegenerative
 CC disorders, developmental abnormalities and foetal deficiencies, blood
 CC disorders, tumours, leukemias, diseases of the immune system, autoimmune
 CC diseases, hepatic and renal disease, lymphomas, inflammation, allergies,
 CC ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,
 CC testostosis, prostate diseases, obesity, disorders involving osteoclasts
 CC such as osteoporosis, arthritis or malignancies, diseases of testes,
 CC lung or thymus, digestive/endocrine disorders, infections and AIDS. The
 CC polypeptides are also useful for identifying their binding partners.
 CC The present sequence represents human secreted protein (see descriptor
 CC line for gene number and clone identification).

XX Sequence 206 AA;

Query Match 88.6%; Score 31; DB 20; Length 206;
 Best Local Similarity 62.5%; Pred. NO. 41;

Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 wxxxxfxg 8
 11 11 1
 Db 192 wkysfsfg 199

RESULT 4
 AAG90500
 ID AAG90500 standard; Protein; 301 AA.

XX AAG90500;

XX 26-SEP-2001 (first entry)

DT C glutamicum protein fragment SEQ ID NO: 4254.

DE Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis.

XX Corynebacterium glutamicum.

OS Epl108790-A2.

PN 20-JUN-2001.

PD 18-DEC-2000; 2000EP-0127688.

XX 16-DEC-1999; 990P-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX (KYOW) KYOWA HAKKO KOGYO KK.

PA Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

DR WPI; 2001-376931/40.
 N-PSDB; AAH65719.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 XX Claim 17; SEQ ID NO: 4254; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from coryneform bacterium, and identifying a homologue of a gene derived
 CC from coryneform bacterium. Coryneform bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.

XX Sequence 301 AA;

Query Match 85.7%; Score 30; DB 22; Length 301;
 Best Local Similarity 62.5%; Pred. NO. 94;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxxxfxg 8
 11 11 1
 Db 254 wklsftg 261

Db 192 wkysfsfg 199

RESULT 2

ID AAB6169 standard; protein; 205 AA.

AC AAB6169;

DT 02-APR-2001 (first entry)

DE Protein of the invention #81.

KW Secreted; transmembrane; gene therapy.

XX Unidentified.

OS W0200078961-A1.

PN 28-DEC-2000.

PD 18-FEB-2000; 2000WO-US04342.

PF 23-JUN-1999; 99US-0141037.

PR 20-JUL-1999; 99US-0144758.

PR 26-JUL-1999; 99US-0145698.

PR 01-SEP-1999; 99WO-US20111.

PR 29-OCT-1999; 99US-0162506.

PR 30-NOV-1999; 99WO-US28313.

PR 02-DEC-1999; 99WO-US28551.

PR 16-DEC-1999; 99WO-US30095.

PR 05-JAN-2000; 2000WO-US00219.

PR 06-JAN-2000; 2000WO-US00376.

XX (GETH) GENENTECH INC.

XX Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;

XX Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;

XX Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;

XX Watanabe CK, Williams PM, Wood WI;

XX WPI; 2001-071395/08.

PT Secreted and transmembrane proteins and nucleic acids designated PRO,

PT useful as hybridization probes, in chromosome and gene mapping and gene

PT therapy -

XX Claim 1; Fig 162; 787pp; English.

XX The present invention relates to secreted and transmembrane proteins.

XX These proteins and the DNA encoding them may be used as hybridization

XX probes, in chromosome and gene mapping and in the generation of

XX anti-sense RNA and DNA. They may also be used to generate either

XX transgenic animals or knockout animals which are in turn useful for

XX development and screening of therapeutically useful reagents.

XX The nucleic acids may also be used in gene therapy.

XX Sequence 205 AA;

Query Match 88.6%; Score 31; DB 22; Length 205;

Best Local Similarity 62.5%; Pred. No. 41;

Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsfsg 8

Db 192 wkysfsfg 199

RESULT 3

AAW8747 standard; Protein; 206 AA.

XX AAW8747;

DT 01-MAR-1999 (first entry)

DE Secreted protein encoded by gene 45 clone HCESF40.

KW Human; secreted protein; fusion protein; gene therapy; protein therapy;

KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;

KW developmental abnormality; foetal deficiency; blood; allergy; renal;

KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;

KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;

KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;

KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;

KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

OS Homo sapiens.

PN W09854963-A2.

PD 10-DEC-1998.

PF 04-JUN-1998; 98WO-US11422.

PR 18-DEC-1997; 97US-0070923.

PR 06-JUN-1997; 97US-0048877.

PR 06-JUN-1997; 97US-0048881.

PR 06-JUN-1997; 97US-0048884.

PR 06-JUN-1997; 97US-0048893.

PR 06-JUN-1997; 97US-0048896.

PR 06-JUN-1997; 97US-0048899.

PR 06-JUN-1997; 97US-0048915.

PR 06-JUN-1997; 97US-0048949.

PR 06-JUN-1997; 97US-0048964.

PR 06-JUN-1997; 97US-0048972.

PR 06-JUN-1997; 97US-0049375.

PR 05-SEP-1997; 97US-0057628.

PR 05-SEP-1997; 97US-0057635.

PR 05-SEP-1997; 97US-0057644.

PR 05-SEP-1997; 97US-0057647.

PR 05-SEP-1997; 97US-0057650.

PR 05-SEP-1997; 97US-0057661.

PR 05-SEP-1997; 97US-0057667.

PR 05-SEP-1997; 97US-0057761.

PR 05-SEP-1997; 97US-0057764.

PR 05-SEP-1997; 97US-0057770.

PR 05-SEP-1997; 97US-0057775.

PR 05-SEP-1997; 97US-0057778.

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC


```
PR 09-SEP-1998: 98US-0099598.
PR 09-SEP-1998: 98US-0099602.
PR 09-SEP-1998: 98US-0099642.
PR 10-SEP-1998: 98US-0099741.
PR 10-SEP-1998: 98US-0099753.
PR 10-SEP-1998: 98US-0099763.
PR 10-SEP-1998: 98US-0099792.
PR 10-SEP-1998: 98US-0099808.
PR 10-SEP-1998: 98US-0099815.
PR 10-SEP-1998: 98US-0099815.
PR 10-SEP-1998: 98US-0099815.
PR 15-SEP-1998: 98US-0100385.
PR 15-SEP-1998: 98US-0100388.
PR 15-SEP-1998: 98US-0100390.
PR 16-SEP-1998: 98US-0100584.
PR 16-SEP-1998: 98US-0100627.
PR 16-SEP-1998: 98US-0100661.
PR 16-SEP-1998: 98US-0100665.
PR 16-SEP-1998: 98US-0100666.
PR 17-SEP-1998: 98US-0100683.
PR 17-SEP-1998: 98US-0100684.
PR 17-SEP-1998: 98US-0100710.
PR 17-SEP-1998: 98US-0100711.
PR 17-SEP-1998: 98US-0100919.
PR 17-SEP-1998: 98US-0100930.
PR 18-SEP-1998: 98US-0100848.
PR 18-SEP-1998: 98US-0100849.
PR 18-SEP-1998: 98US-0101014.
PR 18-SEP-1998: 98US-0101068.
PR 18-SEP-1998: 98US-0101071.
PR 22-SEP-1998: 98US-0101279.
PR 23-SEP-1998: 98US-0101471.
PR 23-SEP-1998: 98US-0101474.
PR 23-SEP-1998: 98US-0101474.
PR 23-SEP-1998: 98US-0101475.
PR 23-SEP-1998: 98US-0101476.
PR 23-SEP-1998: 98US-0101477.
PR 23-SEP-1998: 98US-0101479.
PR 24-SEP-1998: 98US-0101738.
PR 24-SEP-1998: 98US-0101741.
PR 24-SEP-1998: 98US-0101743.
PR 24-SEP-1998: 98US-0101915.
PR 24-SEP-1998: 98US-0101916.
PR 29-SEP-1998: 98US-0102207.
PR 29-SEP-1998: 98US-0102240.
PR 29-SEP-1998: 98US-0102307.
PR 29-SEP-1998: 98US-0102330.
PR 29-SEP-1998: 98US-0102331.
PR 30-SEP-1998: 98US-0102484.
PR 30-SEP-1998: 98US-0102487.
PR 30-SEP-1998: 98US-0102570.
PR 30-SEP-1998: 98US-0102571.
PR 01-OCT-1998: 98US-0102684.
PR 01-OCT-1998: 98US-0102687.
PR 02-OCT-1998: 98US-0102965.
PR 06-OCT-1998: 98US-0103258.
PR 07-OCT-1998: 98US-0103449.
PR 07-OCT-1998: 98US-0103314.
PR 07-OCT-1998: 98US-0103315.
PR 07-OCT-1998: 98US-0103328.
PR 07-OCT-1998: 98US-0103335.
PR 07-OCT-1998: 98US-0103396.
PR 07-OCT-1998: 98US-0103401.
PR 08-OCT-1998: 98US-0103633.
PR 08-OCT-1998: 98US-0103678.
PR 08-OCT-1998: 98US-0103679.
PR 08-OCT-1998: 98US-0103711.
PR 14-OCT-1998: 98US-0104257.
PR 20-OCT-1998: 98US-0104987.
PR 20-OCT-1998: 98US-0105000.
PR 20-OCT-1998: 98US-0105002.
PR 21-OCT-1998: 98US-0105104.
PR 22-OCT-1998: 98US-0105169.

PR 22-OCT-1998: 98US-0105266.
PR 26-OCT-1998: 98US-0105693.
PR 26-OCT-1998: 98US-0105694.
PR 27-OCT-1998: 98US-0105807.
PR 27-OCT-1998: 98US-0105811.
PR 27-OCT-1998: 98US-0105882.
PR 27-OCT-1998: 98US-0106023.
PR 28-OCT-1998: 98US-0106029.
PR 28-OCT-1998: 98US-0106030.
PR 28-OCT-1998: 98US-0106030.
PR 28-OCT-1998: 98US-0106033.
PR 28-OCT-1998: 98US-0106178.
PR 29-OCT-1998: 98US-0106248.
PR 29-OCT-1998: 98US-0106384.
PR 29-OCT-1998: 98US-0108500.
PR 30-OCT-1998: 98US-0106464.
PR 03-NOV-1998: 98US-0106856.
PR 03-NOV-1998: 98US-0106902.
PR 03-NOV-1998: 98US-0106905.
PR 03-NOV-1998: 98US-0106919.
PR 03-NOV-1998: 98US-0106932.
PR 03-NOV-1998: 98US-0106934.
PR 10-NOV-1998: 98US-0107783.
PR 17-NOV-1998: 98US-0108775.
PR 17-NOV-1998: 98US-0108779.
PR 17-NOV-1998: 98US-0108787.
PR 17-NOV-1998: 98US-0108788.
PR 17-NOV-1998: 98US-0108801.
PR 17-NOV-1998: 98US-0108801.
PR 17-NOV-1998: 98US-0108806.
PR 17-NOV-1998: 98US-0108807.
PR 17-NOV-1998: 98US-0108867.
PR 18-NOV-1998: 98US-0108825.
PR 18-NOV-1998: 98US-0108848.
PR 18-NOV-1998: 98US-0108849.
PR 18-NOV-1998: 98US-0108850.
PR 18-NOV-1998: 98US-0108851.
PR 18-NOV-1998: 98US-0108852.
PR 18-NOV-1998: 98US-0108858.
PR 18-NOV-1998: 98US-0108904.

(GETH ) GENENTECH INC.
XX PA Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
XX PT N-PSDB; AAA37102.
XX PT WPL: 2000-237871/20.
XX PT secreted PRO DNA sequences encoding transmembrane, receptor or
XX PT secreted PRO polypeptides, useful for screening of potential peptide or
XX PT small molecule inhibitors of the relevant receptor/ligand interactions
XX PS Claim 12; Fig 162; 773pp; English.
XX XX AAA37022 to AAA37144 encode the new isolated human transmembrane,
XX CC receptor or secreted PRO polypeptides given in AAY9340 to AAY9462. The
XX CC transmembrane and receptor PRO proteins can be used for screening of
XX CC potential peptide or small molecule inhibitors of the relevant
XX CC receptor/ligand interactions. The polypeptides and nucleotide sequences
XX CC encoding then have various industrial applications, including uses as
XX CC pharmaceutical and diagnostic agents. AAA37145 to AAA3730 represent
XX CC PCR primers and hybridisation probes used in the isolation of the PRO
XX CC polypeptides from the present invention.
XX SQ Sequence 205 AA:

Query Match 88.6%; Score 31; DB 21; Length 205;
Best Local Similarity 62.5%; Pred. No. 41;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Oy 1 wxxsfxg 8
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:35:59 ; Search time 103.51 Seconds
(without alignments)
5.725 Million cell updates/sec

Title: 09-185908-1c
Perfect score: 35
Sequence: 1 WXXXSFXG 8

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database :

A_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	88.6	205	21	AAV99420 Human PRO1486 (UNQ
2	31	88.6	205	22	AAB6169 Protein of the inv
3	31	88.6	206	22	AAH88747 Secreted protein e
4	30	85.7	301	22	AAG90500 C glutamic prote
5	30	85.7	345	20	AAV31651 Corynebacterium g1
6	30	85.7	345	20	AAV22645 3-PGDH protein, al
7	30	85.7	530	20	AAV31650 Brevibacterium fla
8	30	85.7	530	20	AAV31649 Brevibacterium fla
9	30	85.7	530	20	AAV22646 Wild type 3-PGDH p
10	30	85.7	530	20	AAV22647 Mutant 3-PGDH prot
11	30	85.7	530	22	AAV91161 C glutamic prote

12	29	82.9	75	21	AAB53886 Human colon cancer
13	28	80.0	55	20	AAV12227 Human 5' EST secre
14	28	80.0	56	22	AAE04286 Human gene 10 enco
15	28	80.0	72	20	AAV32938 Human cerebellin-2
16	28	80.0	83	22	AAV40407 Human polypeptide
17	28	80.0	113	22	AAE04281 Human gene 10 enco
18	28	80.0	114	20	AAV12226 Human 5' EST secre
19	28	80.0	126	22	AAV25829 Human protein sequ
20	28	80.0	126	22	AAV42193 Human polypeptide
21	28	80.0	130	20	AAV01484 Cerebellin protein
22	28	80.0	140	22	AAE04228 Human gene 10 enco
23	28	80.0	155	21	AAE54052 Human pancreatic c
24	28	80.0	224	20	AAV32937 Human cerebellin-2
25	28	80.0	224	22	AAU12214 Human PRO4338 poly
26	28	80.0	230	20	AAV36134 Human secreted pro
27	28	80.0	230	20	AAV36181 Human secreted pro
28	28	80.0	230	21	AAV99378 Human PRO1356 (UNQ
29	28	80.0	230	21	AAV84609 A human membrane a
30	28	80.0	230	22	AAV38857 Human polypeptide
31	28	80.0	230	22	AAU12417 Human PRO1356 poly
32	28	80.0	230	22	AAE04207 Human gene 10 enco
33	28	80.0	230	22	AAE87565 Human PRO1356. Ho
34	28	80.0	230	22	AAE88342 Human membrane or
35	28	80.0	230	22	AAE66127 Protein of the inv
36	28	80.0	260	22	AAV40643 Human polypeptide
37	28	80.0	552	20	AAV29648 Sheep prostaglandi
38	28	80.0	600	10	AAV91008 Prostaglandin endo
39	28	80.0	746	20	AAV29512 Human lung tumour
40	28	80.0	746	21	AAV44461 Human lung tumour
41	28	80.0	951	20	AAV34536 Porphyromonas ging
42	28	80.0	953	20	AAV34403 Porphyromonas ging
43	27	77.1	8	21	AAB06419 Claudin-1 cell adh
44	27	77.1	8	21	AAB06479 Claudin-1 cyclic c
45	27	77.1	8	21	AAB06764 Claudin-6/9 cell a
46	27	77.1	8	21	AAB06916 Claudin-6/9 cyclic
47	27	77.1	8	21	AAB06916 Claudin cell adhes
48	27	77.1	9	21	AAB06420 Claudin-1 cell adh
49	27	77.1	10	21	AAB06436 Claudin-1 cyclic c
50	27	77.1	10	21	AAB06445 Claudin-1 cyclic c

ALIGNMENTS

RESULT 1
ID AAV99420 standard; Protein; 205 AA.
AAV99420;
08-AUG-2000 (first entry)
Human PRO1486 (UNQ755) amino acid sequence SEQ ID NO:287.

Human: PRO polypeptide; membrane bound protein; receptor; diagnosis;
transmembrane; secretion; Immunoadhesion; pharmaceutical; screening.
Homo sapiens.
WO200012708-A2.
09-MAR-2000.
01-SEP-1999; 99WO-US20111.
01-SEP-1998; 98US-0098746.
01-SEP-1998; 98US-0098749.
01-SEP-1998; 98US-0098750.
02-SEP-1998; 98US-0098803.
02-SEP-1998; 98US-0098821.
02-SEP-1998; 98US-0098843.
09-SEP-1998; 98US-0099536.
09-SEP-1998; 98US-0099596.

DR InterPro: IPR000574; Tymo_coat.
DR Pfam: PF00983; Tymo_coat; 1.
KW Coat protein.
FT NON_TER 1
SQ SEQUENCE 175 AA; 18561 MW; 2901894370F4DDCA CRC64;

Query Match 77.1%; Score 27; DB 12; Length 175;
Best Local Similarity 50.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wkxxsfxg 8
|| |
Db 20 WKVTDFTG 27

RESULT 20
O41983 PRELIMINARY; PRT; 177 AA.
AC O41983;
DT 01-JAN-1998 (TRENBLREL. 05, Created)
DT 01-JAN-1998 (TRENBLREL. 05, Last sequence update)
DT 01-JUN-2001 (TRENBLREL. 17, Last annotation update)
DE COAT PROTEIN (FRAGMENT).
OS maize rayado fino virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Marafivirus.
OX NCBI_TaxID=59749;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BOLIVIA A;
RX MEDLINE=98062136; Pubmed=9400964;
RA Hammond R.W., Kogel R., Ramirez P.;
RT "Variability of geographically distinct isolates of maize rayado fino
virus in Latin America";
RL J. Gen. Virol. 78:0-0(0);
DR EMBL; U97717; AAB96562.1; -;
DR InterPro: IPR000574; Tymo_coat.
DR Pfam: PF00983; Tymo_coat; 1.
KW Coat protein.
FT NON_TER 1
SQ SEQUENCE 177 AA; 18661 MW; E91A97A26A575EC1 CRC64;

Query Match 77.1%; Score 27; DB 12; Length 177;
Best Local Similarity 50.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wkxxsfxg 8
|| |
Db 22 WKVTDFTG 29

Search completed: January 14, 2002, 07:39:43
Job time: 954 sec

RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, COLUMBIA;
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
 RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
 RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
 RT "Arabidopsis thaliana chromosome III BAC T16011 genomic sequence.";
 RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
 RN 121
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, COLUMBIA;
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
 RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
 RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
 RT "Arabidopsis thaliana chromosome III PL MZB10 genomic sequence.";
 RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AC010871; AAF07825.1; -;
 DR EMBL: AC009326; AAD56319.1; -;
 DR InterPro: IPR000210; BTB_P02.
 DR InterPro: IPR001005; MYB_DNA_bind.
 DR InterPro: IPR001331; K_tetra.
 DR Pfam: PF02214; K_tetra.1.
 DR PROSITE: PS00037; MYB_1; UNKNOWN_1.
 DR SMART: SM00225; BTB; 1.
 SQ SEQUENCE 460 AA; 49278 MW; 59CCB10EE974649E CRC64;

Query Match 80.0%; Score 28; DB 10; Length 460;
 Best Local Similarity 50.0%; Pred. No. 1.8e+02;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsfxxg 8
 Db 448 WETSSFG 455

RESULT 17
 Q9RPH6 PRELIMINARY; PRT; 1083 AA.
 AC Q9RPH6;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE RECB.
 GN RECB.
 OS Mycobacterium smegmatis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1772;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN=MC2155 (NB2);
 RC MEDLINE=99412429; PubMed=10481025;
 RA Griffin T.J. IV, Parsons L., Leschzner A.E., Devost J.,
 RA Derbyshire K.M., Grindley N.D.F.;
 RT "In vitro transposition of Tn552: a tool for DNA sequencing and
 mutagenesis.";
 RL Nucleic Acids Res. 27:3859-3865(1999).
 DR EMBL: AF157643; AAD46808.1; -;
 DR HSSP: P56255; 2PJR
 DR InterPro: IPR001993; Mitoch_carrier.
 DR InterPro: IPR000212; UVRD_helicase.
 DR Pfam: PF00580; UVRD_helicase.1.
 DR PROSITE: PS00215; MITOCH_CARRIER; UNKNOWN_1.
 SQ SEQUENCE 1083 AA; 117442 MW; A531F5FBA85EB748 CRC64;

Query Match 80.0%; Score 28; DB 2; Length 1083;
 Best Local Similarity 50.0%; Pred. No. 4.3e+02;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 wkxxsfxxg 8
 Db 11-1-1

Db 766 WKRTSYSG 773

RESULT 18
 Q9LRV0 PRELIMINARY; PRT; 1168 AA.
 AC Q9LRV0;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE PRE-MRNA SPLICING FACTOR ATP-DEPENDENT RNA HELICASE-LIKE PROTEIN.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
 RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
 RN 121
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE=20277480; PubMed=10819329;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
 RT features of the regions of 4,504,864 bp covered by sixty pl and TAC
 RT clones.";
 RL DNA Res. 7:131-135(2000).
 DR EMBL: AB028611; BAB01838.1; -;
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002464; DEAD_ATP_helicase.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR003029; SL.
 DR Pfam: PF00575; SL.1.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELIC_C.1.
 DR SMART: SM00316; SL; 1.
 DR PROSITE: PS00690; DEAD_ATP_HELICASE.1.
 KW Helicase.
 SQ SEQUENCE 1168 AA; 134156 MW; B3632DEA7A7690C CRC64;

Query Match 80.0%; Score 28; DB 10; Length 1168;
 Best Local Similarity 50.0%; Pred. No. 4.7e+02;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsfxxg 8
 Db 1002 WKAKNPSG 1009

RESULT 19
 Q9QDK6 PRELIMINARY; PRT; 175 AA.
 ID Q9QDK6;
 AC Q9QDK6;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE COAT PROTEIN (FRAGMENT).
 OS maize rayado fino virus.
 OS maize rayado fino virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Marafivirus.
 OX NCBI_TaxID=59749;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRASIL 17;
 RA Hammond R.W., Bedendo I.P.;
 RT "Molecular confirmation of maize rayado fino virus as the Brazilian
 RT corn streak virus and its present association with 'red stunt' disease
 RT in Brazil.";
 RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF186177; AAD56416.1; -;

OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CE31A;
 RA Martinez-Abarca F., Toro N.;
 RT "Bacterial group II intron mobility into DNA transposition sites in vivo."
 RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AJ242574; CAB43595.1; -
 DR InterPro: IPR001584; Rve.
 DR Pfam: PF00665; Rve; 1.
 SQ SEQUENCE 315 AA; 35780 MW; 414809544109BEE5 CRC64;

Query Match 80.0%; Score 28; DB 2; Length 315;
 Best local Similarity 50.0%; Pred. No. 1.2e+02;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsfxg 8
 || || |
 Db 181 WKTTFTIG 188

RESULT 13
 ID 09R9L5 PRELIMINARY; PRT; 315 AA.
 AC 09R9L5; 09R9L5;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE TRANSPOSASE HOMOLOG.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Plasmid pSym.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BE31CG;
 RA MEDLINE-20035540; PubMed-10564804;
 RA Blondi E.G., Fancelli S., Bazzicalupo M.;
 RT "ISRM10: a new insertion sequence of Sinorhizobium meliloti:
 RT nucleotide sequence and geographic distribution."
 RL FEMS Microbiol. Lett. 181:171-176(1999).
 DR EMBL: AF143444; AAD37358.1; -
 DR InterPro: IPR001584; Rve.
 DR Pfam: PF00665; Rve; 1.
 KW Plasmid.
 SQ SEQUENCE 315 AA; 35878 MW; 1DA5088F3197D140 CRC64;

Query Match 80.0%; Score 28; DB 2; Length 315;
 Best local Similarity 50.0%; Pred. No. 1.2e+02;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsfxg 8
 || || |
 Db 181 WKTTFTIG 188

RESULT 14
 ID 09B0E2 PRELIMINARY; PRT; 412 AA.
 AC 09B0E2;
 DT 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE PORTAL PROTEIN.
 OS Staphylococcus aureus temperate phage phiSLT.
 OC Viruses.
 OX NCBI_TaxID=130478;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Narita S., Kaneko J., Chiba J., Etienne J., Piemont Y., Kamio Y.;
 RT "Phage conversion of Pantoea-Valentine leukocidin (PVL) in
 RT Staphylococcus aureus: molecular analysis of a PVL-converting phage,
 RT phiSLT."
 RL Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AB045978; BAB21733.1; -
 SQ SEQUENCE 412 AA; 47728 MW; 0E36224385DC8A35 CRC64;

Query Match 80.0%; Score 28; DB 9; Length 412;
 Best local Similarity 62.5%; Pred. No. 1.6e+02;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsfxg 8
 || || |
 Db 35 WKNRSPWG 42

RESULT 15
 ID 022715 PRELIMINARY; PRT; 428 AA.
 AC 022715;
 DT 01-JAN-1998 (TREMblrel. 05, Created)
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE PUTATIVE CLATHRIN COAT ASSEMBLY PROTEIN.
 GN P8A5.29.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Pederspiel N.A., Palm C.J., Conway A.B., Kurtz D.B., Conway A.R.,
 RA Au M., Araujo R., Brendel V., Buehler E., Dewar K., Peng J., Kim C.,
 RA Li Y., Oji O., Osborne B.L., Shinn P., Sun H., Tortum M.,
 RA Vysotskaya V.S., Yu G., Ecker J., Theologis A., Davis R.W.;
 RL Submitted (OCT-1997) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AC002922; AAB71967.1; -
 DR Mendel; 26469; Arabid; 3328; 26469.
 DR InterPro: IPR001392; Adap_comp_sub.
 DR Pfam: PF00928; Adap_comp_sub; 1.
 DR PRINTS: PR00314; CLATHRINADPT.
 DR PROSITE: PS00990; CLAT_ADAPTOR_M_1; 1.
 DR PROSITE: PS00991; CLAT_ADAPTOR_M_2; 1.
 SQ SEQUENCE 428 AA; 49032 MW; 1C3C18CF06919C4F CRC64;

Query Match 80.0%; Score 28; DB 10; Length 428;
 Best local Similarity 62.5%; Pred. No. 1.7e+02;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsfxg 8
 || || |
 Db 346 WKIKSPFG 353

RESULT 16
 ID 09S7R7 PRELIMINARY; PRT; 460 AA.
 AC 09S7R7;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE T16O11.1 PROTEIN.
 GN T16O11.1 OR MZB10.6.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;

RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK026222; BAB15398.1; -
 DR InterPro: IPR001073; Clq.
 DR Pfam: PF00386; Clq; 1.
 DR PRINTS: PR00007; COMPLEMENTC1Q.
 DR SMART: SM00110; Clq; 1.
 DR PROSITE: PS01113; Clq; UNKNOWN_1.
 SO SEQUENCE 158 AA; 17625 MW; 47DB10EDD6DC9760 CRC64;

Query Match 80.0%; Score 28; DB 4; Length 158;
 Best Local Similarity 50.0%; Pred. No. 61;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsfxxg 8
 ||:|
 Db 145 WKSTFSG 152

RESULT 9
 ID 041989 PRELIMINARY; PRT; 177 AA.
 AC 041989;
 DT 01-JAN-1998 (Tremblrel. 05, Created)
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DE COAT PROTEIN (FRAGMENT).
 OS maize rayado fino virus.
 OC Viruses: ssRNA positive-strand viruses, no DNA stage; Marafivirus.
 NCBI_TaxID=59745;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COSTA RICA C;
 RX MEDLINE=98062136; PubMed=9400964;
 RR Hammond R.W., Kogel R., Ramirez P.;
 RT "Variability of geographically distinct isolates of maize rayado fino virus in Latin America";
 RL J. Gen. Virol. 78:0-0(0).
 DR EMBL: U97723; AAB96568.1; -
 DR InterPro: IPR000574; Tympo_coat.
 DR Pfam: PF00983; Tympo_coat; 1.
 KW Coat protein.
 FT NON_TER
 SO SEQUENCE 177 AA; 19102 MW; CE4192EF0F105C1 CRC64;

Query Match 80.0%; Score 28; DB 12; Length 177;
 Best Local Similarity 50.0%; Pred. No. 69;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsfxxg 8
 ||:|
 Db 22 WKVTNPTG 29

RESULT 10
 ID 09ET38 PRELIMINARY; PRT; 193 AA.
 AC 09ET38;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DE CLAUDIN-19 (FRAGMENT).
 OS Mus musculus (Mouse).
 OC Mammalia: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
 OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ICR;
 RA Kluichl Y., Morita K., Furuse M., Tsukita S.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC249889; AAF98323.1; -

DR InterPro: IPR001832; Claudin.
 DR InterPro: IPR000729; PMP22.Claudin.
 DR Pfam: PF00822; PMP22.Claudin; 1.
 DR PRINTS: PR01077; CLAUDIN.
 DR PROSITE: PS01346; CLAUDIN; UNKNOWN_1.
 FT NON_TER
 FT NON_TER 193 193
 SO SEQUENCE 193 AA; 20299 MW; 2F2D82DB5FCF0D7F CRC64;

Query Match 80.0%; Score 28; DB 11; Length 193;
 Best Local Similarity 50.0%; Pred. No. 75;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsfxxg 8
 ||:|
 Db 20 WKOSTYAG 27

RESULT 11
 ID 09RRD0 PRELIMINARY; PRT; 206 AA.
 AC 09RRD0;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DE 3-DEHYDROXYLUBIQUINONE-9-3-METHYLTRANSFERASE, PUTATIVE.
 GN DR2562.
 OS Deinococcus radiodurans.
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
 NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RI;
 RX MEDLINE=20036896; PubMed=10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Uitterback T., Zaleski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1";
 RL Science 286:1571-1577(1999).
 DR EMBL: AE002085; AAF12100.1; -
 DR TIGR: DR2562; -
 DR InterPro: IPR001601; Meth-transf.
 DR InterPro: IPR000051; SAM_bind.
 KW Transferase; Methyltransferase; Ubiquinone; Complete proteome.
 SO SEQUENCE 206 AA; 22186 MW; 6F53E1369E12D870 CRC64;

Query Match 80.0%; Score 28; DB 2; Length 206;
 Best Local Similarity 50.0%; Pred. No. 80;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsfxxg 8
 ||:|
 Db 111 WRTASFDG 118

RESULT 12
 ID 09X983 PRELIMINARY; PRT; 315 AA.
 AC 09X983;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DE ORFAB.
 GN ISRM10-2.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

DR Pfam: PF00069; kinase; 1.
DR PRINTS: PR00019; IEURICHRPT.
DR SMART: SM00370; LRR; 16.
DR SMART: SM00220; S_TKC; 1.
DR SMART: SM00219; TYRC; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS01008; PROTEIN_KINASE_ST; UNKNOWN_1.
DR ATP-binding: Kinase; Transferase.
SQ SEQUENCE 895 AA; 98906 MW; 18CD89626834CD06 CRC64;

Query Match
Best Local Similarity 85.7%; Score 30; DB 10; Length 895;
Matches 5; Conservative 62.5%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsfxg 8
|| || |
Db 621 WKMSFAG 628

RESULT 5
ID Q9X982 PRELIMINARY; PRT; 311 AA.
AC Q9X982;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE ORFAB.
OS ISRM10-1.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2011;
RA Martinez-Abarca F., Toro N.;
RT "Bacterial group II intron mobility into DNA transposition sites in vivo";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ242573; CAB43594.1; .
SQ SEQUENCE 311 AA; 34583 MW; 90AF46A01EA80891 CRC64;

Query Match
Best Local Similarity 82.9%; Score 29; DB 2; Length 311;
Matches 4; Conservative 50.0%; Pred. No. 73;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsfxg 8
|| || |
Db 177 WKTTFGTG 184

RESULT 6
ID Q9G147 PRELIMINARY; PRT; 348 AA.
AC Q9G147;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE MAURURSE K (FRAGMENT).
GN MARK.
OS Daphniophyllum sp. Qiu 94162.
OS Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Saxifragales; Daphniophyllaceae; Daphniophyllum.
OX NCBI_TaxID=132526;
RN [1]
RP SEQUENCE FROM N.A.
RA Fishbein M., Hirsch-Jetter C., Solits D.E., Hufford L.;
RT "Phylogeny of Saxifragales (Angiosperms, Eudicots): Analysis of Rapid, Ancient Radiation.";

RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF274612; AAG01302.1; .
DR InterPro: IPR002866; MatK_N.
DR Pfam: PF01824; MatK_N; 1.
KW Chloroplast.
FT NON_TER 348 348
SQ SEQUENCE 348 AA; 41584 MW; F7B26D4E09B67DFE CRC64;

Query Match
Best Local Similarity 82.9%; Score 29; DB 8; Length 348;
Matches 5; Conservative 62.5%; Pred. No. 81;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsfxg 8
|| || |
Db 77 WKQNSFLG 84

RESULT 7
ID O67832 PRELIMINARY; PRT; 392 AA.
AC O67832;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE HYPOTHETICAL 45.3 KDA PROTEIN.
GN AO_2044.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; Pubmed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Anjay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.";
RL Nature 392:353-358(1998).
DR EMBL: AE000769; AAC07799.1; .
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 392 AA; 45252 MW; 4F9FBB04BB5989C CRC64;

Query Match
Best Local Similarity 82.9%; Score 29; DB 2; Length 392;
Matches 5; Conservative 62.5%; Pred. No. 92;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsfxg 8
|| || |
Db 127 WKRSFEG 134

RESULT 8
ID Q9H667 PRELIMINARY; PRT; 158 AA.
AC Q9H667;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CDNA: FLJ22569 FTS, CLONE HS102142.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HUMAN SMALL INTESTINE;
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Ohayashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isono T., Sugano S.;
RT "NEDO human cDNA sequencing project.";


```

RESULT 2
O9ADD0
AC 09ADD0; PRELIMINARY; PRT: 243 AA.
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HYPOTHETICAL. 27.3 KDA PROTEIN.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID:1902;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Saunders D.C., Harris D.;
RA Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
RL 12
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Cordono A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RA Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
RN 13
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RA "A set of ordered cosmids and a detailed genetic and physical map for
RA the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL MOL. MICROBIOL. 21:77-96(1996).
DR EMBL: AL589707; CAC33901.1;
DR Hypothetical protein.
SO SEQUENCE 243 AA; 27350 MW; 819C5DPF736DDE70 CRC64;

Query Match 88.6%; Score 31; DB 2; Length 243;
Best Local Similarity 62.5%; Pred. No. 20;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsfxxg 8
DB 47 WKASFLG 54

RESULT 3
O9VA04 PRELIMINARY; PRT: 619 AA.
AC 09VA04;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CG15560 PROTEIN.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephyroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID:7227;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RA MEDLINE:20196006; PubMed:10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Blundon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

```

```

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtils K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertlera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glöckner A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostlin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kerchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mactel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaes R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003777; AAF57124.1;
DR FLYBase: FBgn0039826; CG15560.
DR InterPro: IPR003986; NIDogen_ext.
DR SMART: SM00539; NIDO.1.
SO SEQUENCE 619 AA; 69911 MW; 08C3D5128BE30D13 CRC64;

Query Match 85.7%; Score 30; DB 5; Length 619;
Best Local Similarity 62.5%; Pred. No. 87;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wkxxsfxxg 8
DB 303 WKVNSFAG 310

RESULT 4
O9FCN6 PRELIMINARY; PRT: 895 AA.
AC 09FCN6;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE RECEPTOR PROTEIN KINASE-LIKE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Euphorbiales; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI."
RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AB025621; BAB09746.1;
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR003592; LRR_out.
DR InterPro: IPR002290; Ser_thr_kin_actsite.
DR InterPro: IPR001245; Tyr_kin.
DR Pfam: PF00560; LRR_11.

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:39:41 ; Search time 112.89 Seconds
(without alignments)
10.366 Million cell updates/sec

Title: 09-185908-1c
Perfect score: 35
Sequence: 1 wkxxsfsg 8

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database :

1: SPTRMBL_17:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phase:*
11: sp_plant:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	88.6	197	11 Q9JHG0	Q9jhg0 mus musculu
2	31	88.6	243	2 Q9ADD0	Q9add0 streptomyce
3	30	85.7	619	5 Q9VAD4	Q9vad4 drosophila
4	30	85.7	895	10 Q9FGN6	Q9fgn6 arabidopsis
5	29	82.9	311	2 Q9X982	Q9x982 rhizobium m
6	29	82.9	348	8 Q9G147	Q9g147 daphniaphyll
7	29	82.9	392	2 Q67832	Q67832 aquilex aeo
8	28	80.0	158	4 Q9H667	Q9h667 homo sapien
9	28	80.0	177	12 Q41989	Q41989 maize rayad
10	28	80.0	193	11 Q9ET38	Q9et38 mus musculu
11	28	80.0	206	2 Q9AR00	Q9ar00 delnococtus
12	28	80.0	315	2 Q9X983	Q9x983 rhizobium m
13	28	80.0	412	9 Q9R915	Q9r915 rhizobium m
14	28	80.0	412	9 Q9B0E2	Q9b0e2 staphylococ
15	28	80.0	428	10 Q22715	Q22715 arabidopsis
16	28	80.0	460	10 Q9S7R7	Q9s7r7 arabidopsis
17	28	80.0	1083	2 Q9RPH6	Q9rph6 mycobacteri
18	28	80.0	1168	10 Q9LRV0	Q9lrvo arabidopsis
19	27	77.1	175	12 Q9QDK6	Q9qdk6 maize rayad

20	27	77.1	177	12	Q41983	Q41983 maize rayad
21	27	77.1	177	12	Q41985	Q41985 maize rayad
22	27	77.1	177	12	Q41986	Q41986 maize rayad
23	27	77.1	177	12	Q41987	Q41987 maize rayad
24	27	77.1	177	12	Q41988	Q41988 maize rayad
25	27	77.1	177	12	Q41990	Q41990 maize rayad
26	27	77.1	177	12	Q41991	Q41991 maize rayad
27	27	77.1	177	12	Q41992	Q41992 maize rayad
28	27	77.1	177	12	Q41993	Q41993 maize rayad
29	27	77.1	177	12	Q41994	Q41994 maize rayad
30	27	77.1	177	12	Q41995	Q41995 maize rayad
31	27	77.1	177	12	Q41996	Q41996 maize rayad
32	27	77.1	177	12	Q9QDK5	Q9qdk5 maize rayad
33	27	77.1	177	12	Q9QDK4	Q9qdk4 maize rayad
34	27	77.1	369	2	Q59399	Q59399 escherichia
35	27	77.1	369	2	Q24687	Q24687 anabaena az
36	27	77.1	369	2	Q9A5D4	Q9a5d4 caulobacter
37	27	77.1	377	5	Q9N1D0	Q9njd0 branchiosto
38	27	77.1	377	5	Q9N1C3	Q9njc3 branchiosto
39	27	77.1	377	5	Q9B334	Q9bj34 branchiosto
40	27	77.1	377	5	Q9B333	Q9bj33 branchiosto
41	27	77.1	378	2	Q9URB0	Q9urb0 neisseria m
42	27	77.1	379	2	Q84731	Q84731 chlamydia t
43	27	77.1	379	2	Q9PLK0	Q9plk0 chlamydia m
44	27	77.1	379	2	Q9K1T9	Q9k1t9 chlamydia p
45	27	77.1	379	10	Q9FND2	Q9fnd2 arabidopsis
46	27	77.1	396	10	Q82014	Q82014 linum usita
47	27	77.1	396	10	Q9SBA2	Q9sba2 linum usita
48	27	77.1	415	2	Q92739	Q92739 chlamydia p
49	27	77.1	432	5	Q9NFP2	Q9nfp2 plasmodium
50	27	77.1	441	10	Q9SLB4	Q9slb4 arabidopsis

ALIGNMENTS

RESULT 1
Q9JHG0 PRELIMINARY; PRT; 197 AA.
AC Q9JHG0;
DF 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DI 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CBLN3.
GN CBLN3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Pang Z., Morgan J.I.;
RT "Cloning and characterization of a novel precerebellin-related gene."
RI Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF218380; AAF32315.1; -;
DR EMBL; AF218379; AAF32314.1; -;
DR MGD; MGI:189286; Cbln3.
DR InterPro; IPR001073; C1q.
DR Pfam; PF00386; C1q; 1.
DR PRINTS; PR00007; COMPLEMENTC1Q.
DR SMART; SM00110; C1Q; 1.
DR PROSITE; PS01113; C1Q; 1.
SO SEQUENCE 197 AA; 21077 MW; DBA8925C9BB1B77 CRC64;

Query Match 88.6%; Score 31; DB 11; Length 197;
Best local Similarity 62.5%; Pred. No. 16;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 wkxxsfsg 8
DB 184 WKYSSFG 191

Search completed: January 14, 2002, 07:40:37
Job time: 503 sec

```
DR InterPro: IPR000066; LHC.
DR Pfam: PF00556; LHC. 1.
DR PROSITE: PS00968; ANTENNA_COMP_ALPHA; 1.
KW Antenna complex; Light-harvesting polypeptide; Transmembrane;
KW Magnesium; Bacteriochlorophyll; Inner membrane.
FT DOMAIN 1 13 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 14 34 POTENTIAL.
FT DOMAIN 35 65 PERIPLASMIC (POTENTIAL).
FT DOMAIN 29 29 AXIAL LIGAND TO THE BACTERIOCHLOROPHYLL
FT METAL 29 29 AXIAL LIGAND TO THE BACTERIOCHLOROPHYLL
FT SEQUENCE 65 AA; 7688 MW; 55A4C306748E3D9A CRC64;

Query Match 77.1%; Score 27; DB 1; Length 65;
Best Local Similarity 50.0%; Pred. No. 12;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsfxx 8
  ||| |
DB 58 WKRTSYDG 65

RESULT 20
CLDY_BRARE
ID CLDY_BRARE STANDARD; PRT; 208 AA.
AC Q9VH91;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-LIKE PROTEIN ZF-A89.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Keen T.J., Inglehearn C.F.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ011789; CAA09777.1; -.
DR InterPro: IPR001832; Claudin.
DR Pfam: PF00822; PMP22_Claudin; 1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN; 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 160 180 POTENTIAL.
FT SEQUENCE 208 AA; 22205 MW; 6A4BB5EBF3CCAB81 CRC64;

Query Match 77.1%; Score 27; DB 1; Length 208;
Best Local Similarity 50.0%; Pred. No. 36;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsfxx 8
  ||| |
DB 30 WKVTAFIG 37
```

DR PROSITE; PS01186; EGF_2; FALSE_NEG.
KW Oxidoreductase; Dioxxygenase; Peroxidase; Glycoprotein;
KW Prostaglandin biosynthesis; Heme; Iron; Signal; Membrane;
FT EGF-like domain; 3D-structure; Transmembrane.
FT SIGNAL 1 24
FT CHAIN 25 600
FT TRANSMEM 74 82
FT TRANSMEM 86 92
FT TRANSMEM 97 105
FT TRANSMEM 108 122
FT DOMAIN 32 70
FT ACT_SITE 207 207
FT ACT_SITE 385 385
FT BINDING 388 388
FT DISULFID 36 47
FT DISULFID 41 57
FT DISULFID 59 69
FT DISULFID 37 159
FT DISULFID 569 575
FT CARBOHYD 68 68
FT CARBOHYD 144 144
FT CARBOHYD 410 410
FT MOD_RES 530 530
FT VARIANT 97 97
FT VARIANT 164 164
FT VARIANT 456 456
FT VARIANT 520 520
FT VARIANT 525 525
FT MUTAGEN 385 385
FT CONFLICT 1 3
FT CONFLICT 5 5
FT CONFLICT 63 90
FT CONFLICT 92 92
FT CONFLICT 193 193
FT CONFLICT 540 540
SQ SEQUENCE 600 AA; 68791 MW; 1B7B659BBA4353A CRC64;

Query Match 80.0%; Score 28; DB 1; Length 600;
Best Local Similarity 50.0%; Pred. No. 57;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsfsg 8
|| :| |
DB 545 WKAFTFG 552

RESULT 18
ID DDX8_ARATH STANDARD; PRT; 1121 AA.
AC Q38953;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE PRE-MRNA SPLICING FACTOR ATP-DEPENDENT RNA HELICASE.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=97086699; PubMed=8932388;
RA Quigley F., Dao P., Cottet A., Mache R.;
RT "Sequence analysis of an 81 kb contig from Arabidopsis thaliana chromosome III";
RL Nucleic Acids Res. 24:4313-4318(1996).
CC !- FUNCTION: COULD ACT LATE IN THE SPLICING OF PRE-MRNA AND MEDIATE
CC THE RELEASE OF THE SPLICED MRNA FROM SPliceosomes (BY SIMILARITY).
CC !- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC !- SIMILARITY: BELONGS TO THE "DEAD" BOX FAMILY HELICASE. DEAH

CC SUBFAMILY. DDX8/PRP22 ORTHOLOG
CC !- SIMILARITY: CONTAINS 1 'S1 MOTIF' DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

DR EMBL; X98130; CAA66825.1; -
DR EMBL; X97970; CAA66613.1; -
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002464; DEAH_ATP_helcse.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00575; S1; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
DR SMART; SM00316; S1; 1.
DR PROSITE; PS00690; DEAH_ATP_HELICASE; 1.
KW Hypothetical protein; mRNA processing; mRNA splicing; Helicase;
KW ATP-binding; Nuclear protein.
FT DOMAIN 167 236 S1 MOTIF.
FT NP_BIND 491 498 ATP (POTENTIAL).
FT SITE 588 591 DEAH BOX.
FT DOMAIN 730 733 POLY-PRO.
SQ SEQUENCE 1121 AA; 128877 MW; C7DF32B80A35641A CRC64;

Query Match 80.0%; Score 28; DB 1; Length 1121;
Best Local Similarity 50.0%; Pred. No. 1e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsfsg 8
|| :| |
DB 955 WKAKNFG 962

RESULT 19
ID LHA2_ECTHL STANDARD; PRT; 65 AA.
AC P80103;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE LIGHT-HARVESTING PROTEIN B800/830/1020, ALPHA-2 CHAIN (EHS-ALPHA-2)
OS Ectothiorhodospira halochloris.
OC Bacteria; Proteobacteria; gamma subdivision; Ectothiorhodospiraceae;
OC Halorhodospira.
OX NCBI_TaxID=1052;
RN [1]
RP SEQUENCE.
RC STRAIN=DSM 1059;
RX MEDLINE=92249336; PubMed=1577009;
RA Wagner-Huber R., Brunisholz R.A., Bissig I., Frank G., Suter F.,
RA Zuber H.;
RT "The primary structure of the antenna polypeptides of
RT Ectothiorhodospira halochloris and Ectothiorhodospira halophila. Four
RT core-type antenna polypeptides in E. halochloris and E. halophila.";
RL Eur. J. Biochem. 205:917-925(1992).
CC !- FUNCTION: ANTENNA COMPLEXES ARE LIGHT-HARVESTING SYSTEMS, WHICH
CC TRANSFER THE EXCITATION ENERGY TO THE REACTION CENTERS.
CC !- SUBUNIT: THE CORE COMPLEX IS FORMED BY DIFFERENT ALPHA AND BETA
CC CHAINS, BINDING BACTERIOCHLOROPHYLL MOLECULES, AND ARRANGED
CC MOST PROBABLY IN TETRAMERIC STRUCTURES DISPOSED AROUND THE
CC REACTION CENTER. THE NON-PIGMENTED GAMMA CHAINS MAY CONSTITUTE
CC ADDITIONAL COMPONENTS.
CC !- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. INNER MEMBRANE.
CC PIR; S23164; S23164.
DR InterPro; IPR002361; Antenna_comp_alpha.

DR PROSITE; PS00991; CIAT_ADAPTOR_M_2; 1.
 KW Coated pits; Phosphorylation.
 FT CONFLICT 214 214 M -> I (IN REF. 1).
 FT CONFLICT 216 216 D -> H (IN REF. 1).
 FT CONFLICT 222 222 N -> K (IN REF. 1).
 FT CONFLICT 433 433 P -> R (IN REF. 1).
 FT CONFLICT 440 440 I -> M (IN REF. 1).
 FT CONFLICT 450 450 MISSING (IN REF. 1).
 SQ SEQUENCE 475 AA; 53873 MW; C6B11153B845921C CRC64;

Query Match 80.0%; Score 28; DB 1; Length 475;
 Best Local Similarity 62.5%; Pred. No. 46;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsfxg 8
 |||||
 Db 382 WKIRSFPG 389

RESULT 17
 PGH1_SHEEP
 ID PGH1_SHEEP STANDARD; PRT; 600 AA.
 AC P05979;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROSTAGLANDIN G/H SYNTHASE 1 PRECURSOR (EC 1.14.99.1) (CYCLOOXYGENASE
 DE -1) (COX-1) (PROSTAGLANDIN-ENDOROXIDE SYNTHASE 1) (PROSTAGLANDIN H2
 DE SYNTHASE 1) (PGH SYNTHASE 1) (PGHS-1) (PHS 1).
 DE PTGS1 OR COX1.
 OS Ovis aries (Sheep).
 GN Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Ovis.
 OC NCBI_TaxID=9940;
 RN [1]
 RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Vesicular gland;
 RX MEDLINE=8814447; PubMed=3125548;
 RA Dewitt D.L., Smith W.L.;
 RT "Primary structure of prostaglandin G/H synthase from sheep vesicular
 RT gland determined from the complementary DNA sequence."
 RL Proc. Natl. Acad. Sci. U.S.A. 85:1412-1416(1988).
 RN [2]
 RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=88196421; PubMed=3129310;
 RA Yokoyama C., Takai T., Tanabe T.;
 RT "Isolation and characterization of the complementary DNA for sheep
 RT deduced from cDNA sequence."
 RL FEBS Lett. 231:347-351(1988).
 RN [3]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=88153641; PubMed=2831188;
 RA Merlie J., Fagan D., Mudd J., Needleman P.;
 RT "Isolation and characterization of the complementary DNA for sheep
 RT cyclooxygenase."
 RL J. Biol. Chem. 263:3550-3553(1988).
 RN [4]
 RN SEQUENCE OF 523-544.
 RX MEDLINE=84024508; PubMed=6414516;
 RA Roth G.J., Machuga E.T., Ozols J.;
 RT "Isolation and covalent structure of the aspirin-modified,
 RT active-site region of prostaglandin synthetase."
 RL Biochemistry 22:4672-4675(1983).
 RN [5]
 RN HEME-BINDING SITE.
 RX MEDLINE=90203007; PubMed=2108169;
 RA Dewitt D.L., El-Harith E.A., Kraemer S.A., Andrews M.J., Yao E.F.,
 RA Armstrong R.L., Smith W.L.;
 RT "The aspirin and heme-binding sites of ovine and murine prostaglandin
 RT endoperoxide synthases.";

RL J. Biol. Chem. 265:5192-5198(1990).
 RN [6]
 RN ACTIVE SITE TYR-385.
 RX MEDLINE=91056037; PubMed=2122967;
 RA Shimokawa T., Kulmacz R.J., Dewitt D.L., Smith W.L.;
 RT "Tyrosine 385 of prostaglandin endoperoxide synthase is required for
 RT cyclooxygenase catalysis."
 RL J. Biol. Chem. 265:20073-20076(1990).
 RN [7]
 RN CARBOHYDRATE-LINKAGE SITES.
 RX MEDLINE=93352648; PubMed=8349699;
 RA Otto J.C., Dewitt D.L., Smith W.L.;
 RT "N-glycosylation of prostaglandin endoperoxide synthases-1 and -2 and
 RT their orientations in the endoplasmic reticulum."
 RL J. Biol. Chem. 268:18234-18242(1993).
 RN [8]
 RN X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS).
 RX MEDLINE=94166877; PubMed=8121489;
 RA Picot D., Loll P.J., Garavito R.M.;
 RT "The X-ray crystal structure of the membrane protein prostaglandin H2
 RT synthase-1."
 RL Nature 367:243-249(1994).
 RN [9]
 RN X-RAY CRYSTALLOGRAPHY (3.4 ANGSTROMS).
 RX MEDLINE=96022982; PubMed=7552725;
 RA Loll P.J., Picot D., Garavito R.M.;
 RT "The structural basis of aspirin activity inferred from the crystal
 RT structure of inactivated prostaglandin H2 synthase."
 RL Nat. Struct. Biol. 2:637-643(1995).
 CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN REGULATING OR PROMOTING
 CC CELL PROLIFERATION IN SOME NORMAL AND NEOPLASTICALLY TRANSFORMED
 CC CELLS.
 CC -!- CATALYTIC ACTIVITY: ARACHIDONATE + AH(2) + 2 O(2) = PROSTAGLANDIN
 CC H2 + A + H(2)O.
 CC -!- PATHWAY: FIRST STEP IN THE FORMATION OF PROSTAGLANDINS AND
 CC THROMBOXANES.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. ENDOPLASMIC RETICULUM
 CC MEMBRANE AND MICROSOMAL MEMBRANE.
 CC -!- MISCELLANEOUS: THIS ENZYME ACTS BOTH AS A DIOXYGENASE AND AS A
 CC PEROXIDASE.
 CC -!- MISCELLANEOUS: THIS ENZYME IS THE TARGET OF NONSTEROIDAL
 CC ANTI-INFLAMMATORY DRUGS SUCH AS ASPIRIN.
 CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE PROSTAGLANDIN G/H SYNTHASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: J03599; AAA31576.1; -
 CC EMBL: M18243; AAA31511.1; -
 CC EMBL: Y00750; CAA68719.1; -
 CC PIR: A20527; A20527.
 CC PIR: A28960; A28960.
 CC PIR: A29947; A29947.
 CC PIR: S00561; S00561.
 CC PDB: 1PRH; 31-MAR-95.
 CC PDB: 1PGE; 11-JAN-97.
 CC PDB: 1PGF; 11-JAN-97.
 CC PDB: 1PGG; 11-JAN-97.
 CC PDB: 1PTH; 11-APR-96.
 CC PDB: 1CQE; 30-JUN-99.
 CC InterPro: IPR002007; Anim_peroxidse.
 CC InterPro: IPR000561; EGF-like.
 CC InterPro: IPR001536; Peroxidse_3.
 CC PRINTS: PR00457; ANPEROXIDASE.
 CC SMART: SM00181; EGF; 1.
 CC PROSITE: PS00022; EGF_1; FALSE_NEG.

FT METAL 175 175 ZINC (CATALYTIC).
FT BINDING 116 116 IMPORTANT FOR FDH ACTIVITY AND ACTIVATION
FT BY RATTY ACIDS (BY SIMILARITY).
SQ SEQUENCE 375 AA; 39669 MW; 0B9760AB77329FE3 CRC64;

Query Match 80.0%; Score 28; DB 1; Length 375;
Best Local Similarity 50.0%; Pred. No. 37;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 wkxxsfxxg 8
I: | | | |
Db 315 WKATAFGG 322

RESULT 15
SPS2_MOUSE STANDARD; PRT; 452 AA.
AC P97364;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SELENIDE WATER DIKINASE 2 (EC 2.7.9.3) (SELENOPHOSPHATE SYNTHETASE 2)
DE (SELENIUM DONOR PROTEIN 2).
GN SPS2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96017645; PubMed=7588067;
RA Guinaraes M.J., Bazan J.F., Zlotnik A., Wiles M.V., Grimaldi J.C.,
RA Lee F., McClanahan T.;
RT "A new approach to the study of haematopoietic development in the
RT yolk sac and embryoid bodies";
RL Development 121:3335-3346(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97140286; PubMed=8986768;
RA Guinaraes M.J., Peterson D., Vicari A., Cocks B.G., Copeland N.G.,
RA Gilbert D.J., Jenkins N.A., Ferrick D.A., Kastelein R., Bazan J.F.,
RA Zlotnik A.;
RT "Identification of a novel sold homolog from eukaryotes, bacteria,
RT and archaea: is there an autoregulatory mechanism in selenocysteine
RT metabolism";
RL Proc. Natl. Acad. Sci. U.S.A. 93:15086-15091(1996).
CC -1- FUNCTION: SYNTHESIZES SELENOPHOSPHATE FROM SELENIDE AND ATP.
CC -1- CATALYTIC ACTIVITY: ATP + SELENIDE + H(2)O = AMP + SELENOPHOSPHATE
CC + PHOSPHATE.
CC -1- COFACTOR: SELENOCYSTEINE. THE ACTIVE-SITE SELENOCYSTEINE IS
CC ENCODED BY THE OPAL CODON, UGA.
CC -1- SIMILARITY: BELONGS TO THE SELENOPHOSPHATE SYNTHETASE 1 FAMILY.
CC CLASS I SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U43285; AAC53024.1; -;
DR MGD; MGI:108388; Sps2.
DR InterPro: IPR000728; AIRS_related.
DR Pfam: PF00586; AIRS; 1.
KW Transferase; Selenium; Selenocysteine; ATP-binding.
FT ACT_SITE 63 63
FT SE_CYS 63 63 POTENTIAL.
FT SITE 66 66
FT IMPORTANT FOR CATALYTIC ACTIVITY (BY
FT SIMILARITY).
FT NP_BIND 322 328
FT ATP (POTENTIAL).

FT DOMAIN 2 9 POLY-ALA.
FT DOMAIN 433 440 POLY-ALA.
SQ SEQUENCE 452 AA; 47786 MW; 9DA6F7250CFE80E4 CRC64;

Query Match 80.0%; Score 28; DB 1; Length 452;
Best Local Similarity 50.0%; Pred. No. 44;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 wkxxsfxxg 8
I: | | | |
Db 52 WRLTSFG 59

RESULT 16
AP54_YEAST STANDARD; PRT; 475 AA.
AC Q00776;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE CLATHRIN COAT ASSEMBLY PROTEIN AP54 (CLATHRIN COAT ASSOCIATED PROTEIN
DE AP54) (GOLGI ADAPTOR AP-1 54 KDA PROTEIN) (HA1 54 KDA SUBUNIT)
DE (CLATHRIN ASSEMBLY PROTEIN COMPLEX 1 MEDIUM CHAIN).
GN APM1 OR YAP54 OR YPL259C OR P0394.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97104180; PubMed=1761056;
RA Nakayama Y., Goebel M., O'Brine G.B., Lemmon S., Pingshang C.E.,
RA Kirchhausen T.;
RT "The medium chains of the mammalian clathrin-associated proteins have
RT a homolog in yeast";
RL Eur. J. Biochem. 202:569-574(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX Messenguy F., Dubois E., Vierendeels F., Scherens B.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-305 FROM N.A.
RA Dueterhoeft A., Floeth M., Fritz M., Hilbert H., Moestl D.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF THE ADAPTOR COMPLEXES WHICH LINK CLATHRIN
CC TO RECEPTORS IN COATED VESICLES. CLATHRIN-ASSOCIATED PROTEIN
CC COMPLEXES ARE BELIEVED TO INTERACT WITH THE CYTOPLASMIC TAILS OF
CC MEMBRANE PROTEINS, LEADING TO THEIR SELECTION AND CONCENTRATION.
CC AP57 IS PROBABLY A SUBUNIT OF THE GOLGI MEMBRANE ADAPTOR.
CC -1- SUBUNIT: ASSEMBLY PROTEIN COMPLEX 1 (AP-1) IS A HETEROTETRAMER
CC COMPOSED OF TWO LARGE CHAINS (GAMMA AND BETA'), A MEDIUM CHAIN
CC (AP54) AND A SMALL CHAIN (AP19).
CC -1- SUBCELLULAR LOCATION: COATED VESICLE.
CC -1- SIMILARITY: BELONGS TO THE ADAPTOR COMPLEXES MEDIUM SUBUNITS
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X60288; CAA42828.1; -;
DR EMBL; Z73615; CAA97989.1; -;
DR PIR; S17028; S17028.
DR SGD; S0006180; APM1.
DR InterPro: IPR001392; Adap_comp_sub.
DR Pfam: PF00928; Adap_comp_sub; 1.
DR PRINTS; PR00314; CLATHRINADPT.
DR PROSITE; PS00990; CLAT_ADAPTOR_M_1; 1.

```

RC TISSUE=Brain;
RX MEDLINE=91203483; PubMed=1850079;
RA Wada C., Ohtani H.;
RT "Molecular cloning of rat cerebellin-like protein cDNA which encodes
a novel membrane-associated glycoprotein."
RL Brain Res. Mol. Brain Res. 9:71-77(1991).
CC -!- FUNCTION: PROBABLY INVOLVED IN SYNAPTIC FUNCTIONS IN THE CNS.
CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.
CC -!- ALTERNATIVE PRODUCTS: TISSUE SPECIFIC ALTERNATIVE SPLICING OCCURS
IN BRAIN, ADRENAL GLAND AND SPLEEN.
CC -!- TISSUE SPECIFICITY: BRAIN, ADRENAL GLAND AND SPLEEN.
CC -!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
DR PIR; A60032; A60032.
DR InterPro; IPR001073; C1q.
DR PRINTS; PR00007; COMPLEMTNC1Q.
DR SMART; SM00110; C1q; 1.
DR PROSITE; PS01113; C1q; 1.
KW Glycoprotein; Membrane; Transmembrane.
FT TRANSMEM 32 49 POTENTIAL.
FT DOMAIN 50 224 EXTRACELLULAR.
FT DOMAIN 86 224 C1Q.
FT SIMILAR 86 103 CEREBELLIN-LIKE.
FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 224 AA; 24022 MW; A8C3ED240CEA53A7 CRC64;

Query Match 80.0%; Score 28; DB 1; Length 224;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxsfsg 8
DB 211 WKYTFSG 218

RESULT 13
CLD2_HUMAN
ID CLD2_HUMAN STANDARD; PRT; 230 AA.
AC P57739;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE CLAUDIN-2.
GN CLDN2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Epithelium;
RA Reinecker H.-C., Sakaguchi T., Golden H.M.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,
RA Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,
RA Yu J., Han L.H.;
RT "Novel human cDNA clone with function of inhibiting cancer cell
growth."
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

```

CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL; AF250558; AAF98151.1; -
DR EMBL; AF177340; AAG17984.1; -
DR InterPro; IPR001832; Claudin.
DR InterPro; IPR000729; PMP22_Claudin.
DR Pfam; PF00822; PMP22_Claudin; 1.
DR PROSITE; PS01346; PMP22_Claudin; 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 163 183 POTENTIAL.
SQ SEQUENCE 230 AA; 24548 MW; 52CA642D4A62B70D CRC64;

Query Match 80.0%; Score 28; DB 1; Length 230;
Best Local Similarity 50.0%; Pred. No. 24;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxsfsg 8
DB 30 WKTSSYVG 37

RESULT 14
ADHH_GADMO
ID ADHH_GADMO STANDARD; PRT; 375 AA.
AC P81600;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ALCOHOL DEHYDROGENASE CLASS III H CHAIN (EC 1.1.1.1) (GLUTATHIONE-
DEPENDENT FORMALDEHYDE DEHYDROGENASE) (EC 1.2.1.1) (FDH).
OS Gadus morhua (Atlantic cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae;
OC Gadus.
OX NCBI_TaxID=8049;
RN [1]
RP SEQUENCE.
RX MEDLINE=97085413; PubMed=8931553;
RA Danielsson O., Shafqat J., Estonius M., El-Ahmad M., Joernvall H.;
RT "Isozyme multiplicity with anomalous dimer patterns in a class III
alcohol dehydrogenase. Effects on the activity and quaternary
structure of residue exchanges at 'non-functional' sites in a native
protein."
RT Biochemistry 35:14561-14568(1996).
RL Biochemistry 35:14561-14568(1996).
CC -!- FUNCTION: CLASS-III ADH IS REMARKABLY INEFFECTIVE IN OXIDIZING
ETHANOL, BUT IT READILY CATALYZES THE OXIDATION OF LONG-CHAIN
PRIMARY ALCOHOLS AND THE OXIDATION OF S-(HYDROXYMETHYL)
GLUTATHIONE.
CC -!- CATALYTIC ACTIVITY: ALCOHOL + NAD(+) = ALDEHYDE OR KETONE + NADH.
CC -!- CATALYTIC ACTIVITY: FORMALDEHYDE + GLUTATHIONE + NAD(+) =
S-FORMYLGLUTATHIONE + NADH.
CC -!- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY.
CC -!- SUBUNIT: HOMODIMER AND HETERODIMER WITH L CHAIN.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
FAMILY. CLASS-III SUBFAMILY.
DR InterPro; IPR002328; ADH_zinc.
DR InterPro; IPR002085; Adh_zinc.
DR Pfam; PF00107; adh_zinc; 1.
DR PROSITE; PS00059; ADH_ZINC; 1.
KW Oxidoreductase; Zinc; NAD; Multigene family; Acetylation.
MOD_RES 1 1 ACETYLATION.
FT METAL 46 46 ZINC (CATALYTIC).
FT METAL 68 68 ZINC (CATALYTIC).
FT METAL 98 98 ZINC (SECOND ATOM).
FT METAL 101 101 ZINC (SECOND ATOM).
FT METAL 104 104 ZINC (SECOND ATOM).
FT METAL 112 112 ZINC (SECOND ATOM).

```


CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; U33003; AAB06701.1; -
DR InterPro: IPR000498; OmpA_tmem.
DR Pfam: PF01389; OmpA_membrane; 1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 213 25 KDA OUTER-MEMBRANE IMMUNOGENIC
FT PROTEIN.
FT
SQ SEQUENCE 213 AA; 23185 MW; 59003CF46F6D0ED3 CRC64;

Query Match 80.0%; Score 28; DB 1; Length 213;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsfxxg 8
||:|
Db 74 WKAGAFAG 81

RESULT 10
OM25_BRUNE STANDARD; PRT; 213 AA.
AC Q45376;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 25 KDA OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
GN OMP25.

OS Brucella neotoma.
OC Bacteris; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29460;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5K33;
RX MEDLINE=96239016; PubMed=8675306;
RA Cloeckaert A., Verger J.M., Grayon M., Zygmunt M.S., Grepinet O.;
RT "Nucleotide sequence and expression of the gene encoding the major
RT 25-kilodalton outer membrane protein of Brucella ovis: Evidence for
RT antigenic shift, compared with other Brucella species, due to a
RT deletion in the gene";
RL Infect. Immun. 64:2047-2055(1996).
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; U39359; AAB36694.1; -
DR InterPro: IPR000498; OmpA_tmem.
DR Pfam: PF01389; OmpA_membrane; 1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 213 25 KDA OUTER-MEMBRANE IMMUNOGENIC
FT PROTEIN.
FT
SQ SEQUENCE 213 AA; 23167 MW; 49EE0F47B784F87 CRC64;

Query Match 80.0%; Score 28; DB 1; Length 213;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsfxxg 8
||:|
Db 74 WKAGAFAG 81

RESULT 11
OM25_BRUSU STANDARD; PRT; 213 AA.
AC Q45689;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 25 KDA OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
GN OMP25.
OS Brucella suis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1330;
RX MEDLINE=96239016; PubMed=8675306;
RA Cloeckaert A., Verger J.M., Grayon M., Zygmunt M.S., Grepinet O.;
RT "Nucleotide sequence and expression of the gene encoding the major
RT 25-kilodalton outer membrane protein of Brucella ovis: Evidence for
RT antigenic shift, compared with other Brucella species, due to a
RT deletion in the gene";
RL Infect. Immun. 64:2047-2055(1996).
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; U39397; AAB36695.1; -
DR InterPro: IPR000498; OmpA_tmem.
DR Pfam: PF01389; OmpA_membrane; 1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 213 25 KDA OUTER-MEMBRANE IMMUNOGENIC
FT PROTEIN.
FT
SQ SEQUENCE 213 AA; 23151 MW; 55F33CF46F6D0ED3 CRC64;

Query Match 80.0%; Score 28; DB 1; Length 213;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsfxxg 8
||:|
Db 74 WKAGAFAG 81

RESULT 12
CERL_RAT STANDARD; PRT; 224 AA.
ID CERL_RAT
AC P98087;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CEREBELLIN-LIKE GLYCOPROTEIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.

FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 201 25 KDA OUTER-MEMBRANE IMMUNOGENIC
FT PROTEIN.
SQ SEQUENCE 201 AA; 21817 MW; 85A4897489A0935B CRC64;

Query Match 80.0%; Score 28; DB 1; Length 201;
Best Local Similarity 50.0%; Pred. No. 21;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxxsfxxg 8
|| :||
Db 74 WKAGAFAG 81

RESULT 7
OM25_BRUAB ID OM25_BRUAB STANDARD; PRT; 213 AA.
AC Q44664;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 25 KDA OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
GN OMP25.
OS Brucella abortus.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=235;
RN SEQUENCE FROM N.A.
RC STRAIN=544S;
RX MEDLINE=95204367; PubMed=7896724;
RA de Wergifosse P., Lintermans P., Limet J.N., Cloeckaert A.;
RT "Cloning and nucleotide sequence of the gene coding for the major 25-kilodalton outer membrane protein of Brucella abortus.";
RL J. Bacteriol. 177:1911-1914(1995).
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; X79284; CAA55872.1; .
CC InterPro; IPR000498; OmpA_tmem.
CC Pfam; PF01389; OmpA_membrane; 1.
CC Antigen; Outer membrane; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 213 25 KDA OUTER-MEMBRANE IMMUNOGENIC
FT PROTEIN.
FT SEQUENCE 213 AA; 23052 MW; 2328515F1F794BC7 CRC64;

Query Match 80.0%; Score 28; DB 1; Length 213;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxxsfxxg 8
|| :||
Db 74 WKAGAFAG 81

RESULT 8
OM25_BRUCA ID OM25_BRUCA STANDARD; PRT; 213 AA.
AC Q45110;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE 25 KDA OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
GN OMP25.
OS Brucella canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=36855;

RN SEQUENCE FROM N.A.
RP STRAIN=RM6/66;
RX MEDLINE=96239016; PubMed=8675306;
RA Cloeckaert A., Verger J.M., Grayon M., Zygmunt M.S., Grepinet O.;
RT "Nucleotide sequence and expression of the gene encoding the major 25-kilodalton outer membrane protein of Brucella ovis: Evidence for antigenic shift, compared with other Brucella species, due to a deletion in the gene.";
RL Infect. Immun. 64:2047-2055(1996).
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; U39358; AAB36692.1; .
CC InterPro; IPR000498; OmpA_tmem.
CC Pfam; PF01389; OmpA_membrane; 1.
CC Antigen; Outer membrane; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 213 25 KDA OUTER-MEMBRANE IMMUNOGENIC
FT PROTEIN.
FT SEQUENCE 213 AA; 23123 MW; D7E40E247A39B9DF CRC64;

Query Match 80.0%; Score 28; DB 1; Length 213;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxxsfxxg 8
|| :||
Db 74 WKAGAFAG 81

RESULT 9
OM25_BRUME ID OM25_BRUME STANDARD; PRT; 213 AA.
AC Q45321;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 25 KDA OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
GN OMP25.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN SEQUENCE FROM N.A.
RP STRAIN=16M;
RX MEDLINE=96239016; PubMed=8675306;
RA Cloeckaert A., Verger J.M., Grayon M., Zygmunt M.S., Grepinet O.;
RT "Nucleotide sequence and expression of the gene encoding the major 25-kilodalton outer membrane protein of Brucella ovis: Evidence for antigenic shift, compared with other Brucella species, due to a deletion in the gene.";
RL Infect. Immun. 64:2047-2055(1996).
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```
DR InterPro; IPR001073; Clq.
DR Pfam; PF00386; Clq; 1.
DR PRINTS; PR00007; COMPLEMTWC1Q.
DR SMART; SM00110; Clq; 1.
DR PROSITE; PS01113; Clq; 1.
KW Synaptosome; Glycoprotein; Membrane; Signal.
FT SIGNAL 1 21
FT PROPEP 22 56
FT PEPTIDE 57 72
FT PROPEP 73 193
FT DOMAIN 55 193
FT CARBOHYD 23 23
FT CARBOHYD 41 41
FT CARBOHYD 79 79
SQ SEQUENCE 193 AA; 21097 MW; D542FC7987E401A5 CRC64;

Query Match 80.0%; Score 28; DB 1; Length 193;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxxsfxg 8
DB 180 WKYSTFSG 187

RESULT 5
CERB_MOUSE
ID CERB_MOUSE STANDARD; PRT; 193 AA.
AC Q9R171; Q9QV75; P28655;
DT 01-DEC-1992 (Rel. 24, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CEREBELLIN PRECURSOR (CEREBELLIN) (BRAIN PROTEIN D3).
CS CBLN1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP TISSUE=Brain;
RC MEDLINE=95182808; PubMed=7877445;
RA Kavety B., Jenkins N.A., Fletcher C.F., Copeland N.G., Morgan J.I.;
RT "Genomic structure and mapping of precerebellin and a precerebellin-
related gene.";
RL Brain Res. Mol. Brain Res. 27:152-156(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129; TISSUE=Cerebellum;
RX MEDLINE=99058013; PubMed=9838062;
RA Kavety B., Morgan J.I.;
RT "Characterization of transcript processing of the gene encoding
precerebellin-1.";
RL Brain Res. Mol. Brain Res. 63:98-104(1998).
RN [3]
RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Brain;
RA Kato K.;
RT "A collection of cDNA clones with specific expression patterns in
mouse brain.";
RL Eur. J. Neurosci. 2:704-711(1991).
CC -1- FUNCTION: CEREBELLIN EXERTS NEUROMODULATORY FUNCTIONS. DIRECTLY
STIMULATES NOREPINEPHRINE RELEASE VIA THE ADENYLATE CYCLASE/PKA-
DEPENDENT SIGNALING PATHWAY; AND INDIRECTLY ENHANCES
ADRENOCORTICAL SECRETION IN VIVO, THROUGH A PARACRINE MECHANISM
INVOLVING MEDULLARY CATECHOLAMINE RELEASE (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: IN THE PURKINE CELLS POSTSYNAPTIC STRUCTURES.
CC -1- SIMILARITY: CONTAINS 1 ClQ DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF164680; AAD47280.1; -
CC EMBL; X61448; CAA43688.1; ALT_SEQ.
CC PIR; SI8662; SI8662.
CC MGD; MGI:88281; Cbln1.
CC InterPro; IPR001073; Clq.
CC Pfam; PF00386; Clq; 1.
CC SMART; SM00110; Clq; 1.
CC PROSITE; PS01113; Clq; 1.
KW Synaptosome; Glycoprotein; Membrane; Signal.
FT SIGNAL 1 21
FT PROPEP 22 56
FT PEPTIDE 57 72
FT PROPEP 73 193
FT DOMAIN 55 193
FT CARBOHYD 23 23
FT CARBOHYD 79 79
FT CONFLICT 5 5
SQ SEQUENCE 193 AA; 21113 MW; A23C796C7D11BE5F CRC64;

Query Match 80.0%; Score 28; DB 1; Length 193;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxxsfxg 8
DB 180 WKYSTFSG 187

RESULT 6
OM25_BRUOV
ID OM25_BRUOV STANDARD; PRT; 201 AA.
AC Q45335;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 25 KDA OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
CS OMP25.
OS Brucella ovis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=236;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=63/290;
RX MEDLINE=96239016; PubMed=8675306;
RA Cloeckaert A., Verger J.M., Grayon M., Zygmunt M.S., Greplinet O.;
RT "Nucleotide sequence and expression of the gene encoding the major
25-kilodalton outer membrane protein of Brucella ovis: Evidence for
antigenic shift, compared with other Brucella species, due to a
deletion in the gene.";
RL Infect. Immun. 64:2047-2055(1996).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY. HAS A C-TERMINAL
DELETION COMPARED TO THAT OF OTHER BRUCELLA SPECIES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; U33004; AAB06702.1; -
CC InterPro; IPR000498; OmpA_tmem.
CC Pfam; PF01389; OmpA_membrane; 1.
KW Antigen; Outer membrane; Signal.
```



```
DR Pfam; PF01842; ACT; 1.
DR PROSITE; PS00065; D_2-HYDROXYACID_DH_1; 1.
DR PROSITE; PS00670; D_2-HYDROXYACID_DH_2; 1.
DR PROSITE; PS00671; D_2-HYDROXYACID_DH_3; 1.
KW Serine biosynthesis; Oxidoreductase; NAD; Complete proteome.
FT ACT_SITE 232 232 SUBSTRATE-BINDING (BY SIMILARITY).
FT ACT_SITE 261 261 BY SIMILARITY.
FT ACT_SITE 279 279 BY SIMILARITY.
SQ SEQUENCE 528 AA; 54469 MW; 1A6DC60F9FB71222 CRC64;

Query Match      88.6%; Score 31; DB 1; Length 528;
Best Local Similarity 62.5%; Pred. No. 11;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsfxx 8
   |||||
DB 129 WKRSSFSG 136

RESULT 2
SERA_MYCTU
ID SERA_MYCTU STANDARD; PRT; 528 AA.
AC 053243;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE D-3-PHOSPHOGLYCERATE DEHYDROGENASE (EC 1.1.1.95) (PGDH).
GN SERA OR RV2996C OR MT3074 OR MTV012.10.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers R.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: 3-PHOSPHOGLYCERATE + NAD(+) -
CC 3-PHOSPHOHYDROXYPYRUVATE + NADH.
CC -1- PATHWAY: FIRST COMMITTED STEP IN THE "PHOSPHORYLATED" PATHWAY
CC OF L-SERINE BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE D-ISOMER SPECIFIC 2-HYDROXYACID
CC DEHYDROGENASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL; AL021287; CAAL6081.1; -.
DR EMBL; AE007127; AAK47403.1; -.
DR TIGR; MT3074; -.
DR TubercuList; RV2996c; -.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR002162; D_2_hydroxyacid_DH.
DR InterPro; IPR000205; NAD_binding.
DR Pfam; PF00389; 2-Hacid_DH; 1.
DR Pfam; PF01842; ACT; 1.
DR PROSITE; PS00065; D_2-HYDROXYACID_DH_1; 1.
DR PROSITE; PS00670; D_2-HYDROXYACID_DH_2; 1.
DR PROSITE; PS00671; D_2-HYDROXYACID_DH_3; 1.
KW Serine biosynthesis; Oxidoreductase; NAD; Complete proteome.
FT ACT_SITE 232 232 SUBSTRATE-BINDING (BY SIMILARITY).
FT ACT_SITE 261 261 BY SIMILARITY.
FT ACT_SITE 279 279 BY SIMILARITY.
SQ SEQUENCE 528 AA; 54554 MW; 3B5696AAFD82A901 CRC64;

Query Match      88.6%; Score 31; DB 1; Length 528;
Best Local Similarity 62.5%; Pred. No. 11;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wkxxsfxx 8
   |||||
DB 129 WKRSSFSG 136

RESULT 3
KROS_HUMAN
ID KROS_HUMAN STANDARD; PRT; 2347 AA.
AC P08922; Q15368;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PROTO-ONCOGENE TYROSINE-PROTEIN KINASE ROS PRECURSOR (EC 2.7.1.112).
GN ROS1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90280463; PubMed=2352949;
RA Birchmeier C., O'Neill K., Riggs M., Wigler M.;
RT "Characterization of ROS1 cDNA from a human glioblastoma cell line.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4799-4803(1990).
RN [2]
RP SEQUENCE OF 1790-2259 FROM N.A.
RX MEDLINE=87064611; PubMed=3023956;
RA Matsushime H., Wang L.-H., Shibuya M.;
RT "Human c-ros-1 gene homologous to the v-ros sequence of UR2 sarcoma
RT virus encodes for a transmembrane receptorlike molecule.";
RL Mol. Cell. Biol. 6:3000-3004(1986).
RN [3]
RP SEQUENCE OF 1854-2245 FROM N.A.
RX MEDLINE=87064625; PubMed=3785223;
RA Birchmeier C., Birnbaum D., Waechter G., Fasano O., Wigler M.;
RT "Characterization of an activated human ros gene.";
RL Mol. Cell. Biol. 6:3109-3116(1986).
CC -1- FACTOR: THIS IS A PROBABLY A CELL GROWTH OR DIFFERENTIATION
CC FACTOR RECEPTOR WITH A TYROSINE-PROTEIN KINASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE -> ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE INSULIN RECEPTOR FAMILY OF TYROSINE-
CC PROTEIN KINASES. SEVENLESS SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:40:37 ; Search time 37.71 Seconds
(without alignments)
7.778 Million cell updates/sec

Title: 09-185908-1C
Perfect score: 35
Sequence: 1 wkxxsfxxg 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	88.6	528	1 SERA_MYCLE	O33116 mycobacteri
2	31	88.6	528	1 SERA_MYCTU	O53243 mycobacteri
3	29	82.9	2347	1 KROS_HUMAN	P08922 homo sapien
4	28	80.0	193	1 CERB_HUMAN	P23435 homo sapien
5	28	80.0	193	1 CERB_MOUSE	Q9r171 mus musculu
6	28	80.0	201	1 OM25_BRUOV	Q45335 bruceella ov
7	28	80.0	213	1 OM25_BRUAB	Q44664 bruceella ab
8	28	80.0	213	1 OM25_BRUCA	Q45110 bruceella ca
9	28	80.0	213	1 OM25_BRUME	Q45321 bruceella me
10	28	80.0	213	1 OM25_BRUNE	Q45326 bruceella ne
11	28	80.0	213	1 OM25_BRUSU	Q45689 bruceella su
12	28	80.0	224	1 CERL_RAT	P98087 rattus norv
13	28	80.0	230	1 CLD2_HUMAN	P57739 homo sapien
14	28	80.0	375	1 ADHX_GADMO	P81600 gadus morhu
15	28	80.0	452	1 SP52_MOUSE	P97364 mus musculu
16	28	80.0	475	1 AF54_YEAST	Q00776 saccharomyc
17	28	80.0	600	1 PGH1_SHEEP	P05979 ovis aries
18	28	80.0	1121	1 DDX8_ARATH	Q38953 arabidopsis
19	27	77.1	65	1 LHA2_ECTHL	P80103 ectothiorho
20	27	77.1	208	1 CLD2_BRARE	Q9yh91 brachydanio
21	27	77.1	211	1 CLD1_MOUSE	O88551 mus musculu
22	27	77.1	211	1 CLD1_RAT	P56745 rattus norv
23	27	77.1	217	1 CLD9_HUMAN	O95484 homo sapien
24	27	77.1	217	1 CLD9_MOUSE	Q92087 mus musculu
25	27	77.1	219	1 CLD6_MOUSE	Q92362 mus musculu
26	27	77.1	220	1 CLD6_HUMAN	P56747 homo sapien
27	27	77.1	269	1 ADH3_ECOLI	P25437 escherichia
28	27	77.1	369	1 ADH3_PASPI	P39450 pasteurella
29	27	77.1	373	1 ADHX_HORSE	P19854 equus caball
30	27	77.1	373	1 ADHX_HUMAN	P11766 homo sapien
31	27	77.1	373	1 ADHX_MOUSE	P28474 mus musculu
32	27	77.1	373	1 ADHX_RABIT	O19053 oryctolagus
33	27	77.1	373	1 ADHX_RAT	P12711 rattus norv

34	27	77.1	373	1 ADHX_UROHA	P80467 uromastyx h
35	27	77.1	375	1 ADHL_GADMO	P81601 gadus morhu
36	27	77.1	375	1 FADH_PARDE	P45382 paracoccus
37	27	77.1	376	1 ADHI_RHOSH	P72324 rhodobacter
38	27	77.1	376	1 ADHX_MYXGL	P80360 myxine glut
39	27	77.1	376	1 ADHX_SPAAU	P79896 sparus auro
40	27	77.1	378	1 ADHX_HAEIN	P44557 haemophilus
41	27	77.1	378	1 ADHX_DROME	P46415 drosophilla
42	27	77.1	378	1 ADHX_OCTVU	P81431 octopus vul
43	27	77.1	378	1 ADHX_PEA	P80572 pisum sativ
44	27	77.1	379	1 ADHX_ARATH	Q96533 arabidopsis
45	27	77.1	381	1 ADHX_MAIZE	P93629 zea mays (m
46	27	77.1	381	1 ADHX_ORYSA	P93436 oryza sativ
47	27	77.1	381	1 FADH_CANMA	Q06099 candida mal
48	27	77.1	384	1 ADHX_CAEEL	Q17335 caenorhabdi
49	27	77.1	386	1 FADH_YEAST	P32771 saccharomyc
50	27	77.1	396	1 STAD_LINUS	P32062 linum usita

ALIGNMENTS

RESULT 1
SERA_MYCLE
ID SERA_MYCLE STANDARD; PRT; 528 AA.
AC O33116;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE D-3-PHOSPHOGLYCERATE DEHYDROGENASE (EC 1.1.1.95) (PGDH).
GN SERA OR ML1692 OR MLCB637.25.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Fellwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.-A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -!- CATALYTIC ACTIVITY: 3-PHOSPHOGLYCERATE + NAD(+) =
CC 3-PHOSPHOHYDROXYPYRUVATE + NADH.
CC -!- PATHWAY: FIRST COMMITTED STEP IN THE "PHOSPHORYLATED" PATHWAY
CC OF L-SERINE BIOSYNTHESIS.
CC -!- SIMILARITY: BELONGS TO THE D-ISOMER SPECIFIC 2-HYDROXYACID
CC DEHYDROGENASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 299263; CAB16440.1; -;
CC EMBL; AL583923; CAC30645.1; -;
CC HSSP; P01542; 1CCN.
CC Leproma; ML1692;-
CC InterPro; IPR002912; ACT.
CC InterPro; IPR002162; D_2_hydroxyacid_DH.
CC InterPro; IPR000205; NAD_binding.
CC Pfam; PF00389; 2-Hacid_DH; 1.

alcohol dehydrogenase (EC 1.1.1.1) class III - horse
 N:Contains: formaldehyde dehydrogenase (glutathione) (EC 1.2.1.1)
 C:Species: Equus caballus (domestic horse)
 C:Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 01-May-1998
 C:Accession: A33419
 R:Kaiser, R.; Holmquist, B.; Vallee, B.L.; Joernvall, H.
 Biochemistry 28, 8432-8438, 1989
 A:Title: Characteristics of mammalian class III alcohol dehydrogenases, an enzyme less
 A:Reference number: A33419; MUID:90105360
 A:Accession: A33419
 A:Molecule type: protein
 A:Residues: 1-373 <KAI>
 C:Comment: Class III alcohol dehydrogenases are pyrazole-insensitive, are not very active
 C:Function: homodimer; does not form heterodimers with the class I alcohol dehydrogenases
 C:Function: <ADH>
 A:Description: catalyzes the oxidation of primary and secondary alcohols to aldehydes and
 C:Function: <FDH>
 A:Description: catalyzes the oxidation by NAD+ of formaldehyde and glutathione to S-form
 C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
 C:Keywords: acetylated amino end; alcohol metabolism; homodimer; metalloprotein; NAD; ox
 F:1-373/Product: alcohol dehydrogenase chi chain #status experimental <MAT>
 F:29-364/Domain: long-chain alcohol dehydrogenase homology <LADH>
 F:193-222/Region: beta-alpha-beta NAD nucleotide-binding fold
 F:1/Modified site: acetylated amino end (Ser) #status experimental
 F:44,66,173/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
 F:96,99,102,110/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match 77.1%; Score 27; DB 1; Length 373;
 Best Local Similarity 50.0%; Pred. No. 1.2e+02;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 wkxxsfxx 8
 || : ||
 Db 313 WKGTAFGG 320

RESULT 19
 DERTA
 N:Alternate names: alcohol dehydrogenase chi chain; class III alcohol dehydrogenase
 N:Contains: formaldehyde dehydrogenase (glutathione) (EC 1.2.1.1)
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 01-May-1998
 C:Accession: S00331; S02619; S06633
 R:Julia, P.; Pares, X.; Joernvall, H.
 Eur. J. Biochem. 172, 73-83, 1988
 A:Title: Rat liver alcohol dehydrogenase of class III. Primary structure, functional con
 A:Reference number: S00331; MUID:88152004
 A:Accession: S00331
 A:Molecule type: protein
 A:Residues: 1-373 <JUL>
 R:Fairwell, T.; Julia, P.; Kaiser, R.; Holmquist, B.; Pares, X.; Vallee, B.L.; Joernvall
 FEBS Lett. 222, 99-103, 1987
 A:Title: Acetylated N-terminal structures of class III alcohol dehydrogenases. Differenc
 A:Reference number: S02617; MUID:88005160
 A:Accession: S02619
 A:Molecule type: protein
 A:Residues: 1-6 <FAI>
 R:Koivusalo, M.; Baumann, M.; Uotila, L.
 FEBS Lett. 257, 105-109, 1989
 A:Title: Evidence for the identity of glutathione-dependent formaldehyde dehydrogenase a
 A:Reference number: S06633; MUID:90033321
 A:Accession: S06633
 A>Status: preliminary
 A:Molecule type: protein
 A:Residues: 9-25;84-95, 'X',97-98, 'X',187,188,189-193, 'X',195-198;357-365 <KOI>
 A:Experimental source: strain Wistar
 C:Comment: Class III alcohol dehydrogenases are pyrazole-insensitive, are not very active
 C:Genetics:
 A:Gene: ADH-2
 C:Complex: homodimer; does not form heterodimers with the class I alcohol dehydrogenase
 C:Function: <ADH>

A:Description: catalyzes the oxidation of primary and secondary alcohols to aldehy
 C:Function: <FDH>
 A:Description: catalyzes the oxidation by NAD+ of formaldehyde and glutathione to
 C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
 C:Keywords: acetylated amino end; alcohol metabolism; homodimer; metalloprotein;
 F:1-373/Product: alcohol dehydrogenase 2 #status experimental <MAT>
 F:29-364/Domain: long-chain alcohol dehydrogenase homology <LADH>
 F:193-222/Region: beta-alpha-beta NAD nucleotide-binding fold
 F:1/Modified site: acetylated amino end (Ala) #status experimental
 F:44,66,173/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
 F:96,99,102,110/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match 77.1%; Score 27; DB 1; Length 373;
 Best Local Similarity 50.0%; Pred. No. 1.2e+02;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 wkxxsfxx 8
 || : ||
 Db 313 WKGTAFGG 320

RESULT 20
 S68061
 A:Title: Alcohol dehydrogenase (EC 1.1.1.1) class III - Indian spiny-tailed lizard
 N:Contains: formaldehyde dehydrogenase (glutathione) (EC 1.2.1.1)
 C:Species: Uromastyx hardwickii (Indian spiny-tailed lizard)
 C:Date: 19-Mar-1997 #sequence_revision 29-Aug-1997 #text_change 01-May-1998
 C:Accession: S68061; S66194
 R:Hjelmqvist, L.; Shafqat, J.; Siddiqui, A.R.; Joernvall, H.
 FEBS Lett. 373, 212-216, 1995
 A:Title: Alcohol dehydrogenase of class III: consistent patterns of structural an
 A:Reference number: S68061; MUID:96033975
 A:Accession: S68061
 A:Molecule type: protein
 A:Residues: 1-373 <HJE>
 R:Hjelmqvist, L.; Hackett, M.; Shafqat, J.; Danielsson, O.; Iida, J.; Hendrickson
 FEBS Lett. 367, 237-240, 1995
 A:Title: Multiplicity of N-terminal structures of medium-chain alcohol dehydrogen
 nzyme.

A:Reference number: S66191; MUID:95331382
 A:Accession: S66194
 A:Molecule type: protein
 A:Residues: 9-14 <HJM>
 C:Comment: Class III alcohol dehydrogenases are pyrazole-insensitive, are not ver
 C:Complex: homodimer
 C:Function: <ADH>
 A:Description: catalyzes the oxidation of primary and secondary alcohols to aldeh
 C:Function: <FDH>
 A:Description: catalyzes the oxidation by NAD+ of formaldehyde and glutathione to
 C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
 C:Keywords: acetylated amino end; alcohol metabolism; dimer; metalloprotein; NAD;
 F:1-373/Product: alcohol dehydrogenase class III #status experimental <MAT>
 F:29-364/Domain: long-chain alcohol dehydrogenase homology <LADH>
 F:193-222/Region: beta-alpha-beta NAD nucleotide-binding fold
 F:1/Modified site: acetylated amino end (Ala) #status experimental
 F:44,66,173/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
 F:96,99,102,110/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match 77.1%; Score 27; DB 1; Length 373;
 Best Local Similarity 50.0%; Pred. No. 1.2e+02;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 wkxxsfxx 8
 || : ||
 Db 313 WKGTAFGG 320

Search completed: January 14, 2002, 07:37:27
 Job time: 897 sec

F:255-257/281-283/Region: S1 specificity crevice #status predicted
F:409-513/Domain: carboxyl-terminal propeptide #status predicted <CPR>
F:166,197,349/Active site: Asp, His, Ser #status predicted

Query Match 77.8%; Score 28; DB 1; Length 513;
Best Local Similarity 50.0%; Pred. No. 85;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 wrxxsyxg 8
|||
Db 495 WRIVAYSG 502

RESULT 16
S62919
probable membrane protein YNL008c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein N2874
C:Species: Saccharomyces cerevisiae
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Dec-1999
C:Accession: S62919; S62920
R:Dolignon, F.; Crouzet, M.
submitted to the Protein Sequence Database, April 1996
A:Reference number: S62916
A:Accession: S62919
A:Molecule type: DNA
A:Residues: 1-370 <DOI>
A:Cross-references: EMBL:Z71284; GSPDB:GN00014; MIPS:YNL008c
A:Experimental source: strain S288C
R:Andre, B.; Iraqui Housaini, I.; Urrestazu, L.A.; Vissers, S.
submitted to the Protein Sequence Database, April 1996
A:Reference number: S62920
A:Accession: S62920
A:Molecule type: DNA
A:Residues: 103-669 <AND>
A:Cross-references: EMBL:Z71284; GSPDB:GN00014; MIPS:YNL008c
A:Experimental source: strain S288C
C:Genetics:
A:Gene: MIPS:YNL008c
A:Map position: 14L
C:Superfamily: probable membrane protein YMR119w
C:Keywords: transmembrane protein
F:82-98/Domain: transmembrane #status predicted <TM1>
F:117-133/Domain: transmembrane #status predicted <TM2>
F:153-169/Domain: transmembrane #status predicted <TM3>
F:211-227/Domain: transmembrane #status predicted <TM4>
F:275-291/Domain: transmembrane #status predicted <TM5>

Query Match 77.8%; Score 28; DB 1; Length 669;
Best Local Similarity 50.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 wrxxsyxg 8
|||
Db 310 WRSNNYNG 317

RESULT 17
T06088
hypothetical protein T9A14.170 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 22-Oct-1999
C:Accession: T06088
R:Hevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.
submitted to the Protein Sequence Database, March 1999
A:Reference number: Z15184
A:Accession: T06088
A:Molecule type: DNA
A:Residues: 1-700 <BEV>
A:Cross-references: EMBL:AL035656; GSPDB:GN00062; ATSP:T9A14.170
A:Experimental source: cultivar Columbia; BAC clone T9A14
C:Genetics:

A:Gene: ATSP:T9A14.170
A:Map position: 4
A:Introns: 102/3; 139/3; 215/3; 298/1; 386/3; 477/3; 571/3; 612/1; 668/3

Query Match 77.8%; Score 28; DB 2; Length 700;
Best Local Similarity 62.5%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 wrxxsyxg 8
|||
Db 645 WRPPSYFG 652

RESULT 18
H71338
conserved hypothetical protein TP0325 - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 21-Jan-2000
C:Accession: H71338
R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G
ron, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A71250; MUID:98332770
A:Accession: H71338
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-968 <COL>
A:Cross-references: GB:AE001212; GB:AE000520; NID:G3322597; PIDN:AAC65312.1; PID:G332
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP0325
C:Superfamily: syphilis spirochete conserved hypothetical protein TP0325

Query Match 77.8%; Score 28; DB 2; Length 988;
Best Local Similarity 50.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxsyxg 8
|||
Db 441 WRATYEG 448

RESULT 19
T18212
parasporal crystal protein cry5Ab1 - Bacillus thuringiensis (fragment)
C:Species: Bacillus thuringiensis
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C:Accession: T18212
R:Narva, K.E.; Payne, J.M.; Schwab, G.E.; Hickie, L.A.; Galasen, T.; Sick, A.J.
submitted to the EMBL Data Library, July 1999
A:Description: Novel Bacillus thuringiensis microbes active against nematodes, and ge
A:Reference number: Z18829
A:Accession: T18212
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1289 <NAR>
A:Cross-references: EMBL:L07026; NID:G142769; PID:G142770; PIDN:AAA67693.1
C:Genetics:
A:Gene: CryVA(b)
C:Keywords: delta-endotoxin

Query Match 77.8%; Score 28; DB 2; Length 1289;
Best Local Similarity 50.0%; Pred. No. 2.1e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 wrxxsyxg 8
|||
Db 382 WRADYGG 389

RESULT 20

Q4ADE2

early E4 11K protein - human adenovirus 2
C:Species: Mastadenovirus h2 (human adenovirus 2)

A:Note: host Homo sapiens (man)

C:Date: 02-Apr-1982 #sequence_revision 02-Apr-1982 #text_change 04-Mar-1994

C:Accession: A03807

R:Herisse, J.; Rigole, M.; Dupont de Dinechin, S.; Galibert, F.

Nucleic Acids Res. 9, 4023-4042, 1981

A:Title: Nucleotide sequence of adenovirus 2 DNA fragment encoding for the carboxylic re

A:Reference number: A93733; MUID:82059444

A:Accession: A03807

A:Molecule type: DNA

A:Residues: 1-116 <HER>

A:Note: this protein was assigned by correlating EM data and S1 digestion studies

C:Genetics:

A:Map position: 96.0-97.0

C:Superfamily: adenovirus early E4 11K protein

C:Keywords: early protein

Query Match

Best Local Similarity 75.0%; Score 27; DB 1; Length 116;

Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 WXXSYXG 8

11 : 1 1

Db 35 WRDENYLG 42

Search completed: January 14, 2002, 07:37:29
Job time: 899 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:40:38 ; Search time 37.71 Seconds

(without alignments)
7.778 Million cell updates/sec

Title: 09-185908-1e
Perfect score: 36
Sequence: 1 vrxsxyxg 8

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database: SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	88.9	230	1	CLD2_MOUSE
2	32	88.9	1180	1	EX5B_ECOLI
3	31	86.1	211	1	CLD1_HUMAN
4	31	86.1	551	1	YVD3_CAEEL
5	30	83.3	159	1	VAT_CAMVP
6	30	83.3	261	1	ZNUB_ECOLI
7	29	80.6	230	1	CLD2_HUMAN
8	29	80.6	454	1	AOFH_MYCTU
9	29	80.6	822	1	XFP_LACIA
10	28	77.8	65	1	LMA2_ECTHL
11	28	77.8	168	1	GOX2_SULAC
12	28	77.8	191	1	CLD7_RAT
13	28	77.8	211	1	CLD1_MOUSE
14	28	77.8	211	1	CLD1_RAT
15	28	77.8	211	1	CLD7_HUMAN
16	28	77.8	211	1	CLD2_MOUSE
17	28	77.8	452	1	SPS2_MOUSE
18	28	77.8	513	1	AOL1_THESQ
19	28	77.8	669	1	YNAB_YEAST
20	28	77.8	934	1	HMDH_PICJA
21	28	77.8	1289	1	C5AB_BACUD
22	27	75.0	99	1	VAT_CAMVW
23	27	75.0	116	1	EA11_ADEB2
24	27	75.0	116	1	EA11_ADEB5
25	27	75.0	159	1	VAT_CAMVC
26	27	75.0	159	1	VAT_CAMVE
27	27	75.0	159	1	VAT_CAMVE
28	27	75.0	159	1	VAT_CAMV
29	27	75.0	159	1	VAT_CAMV
30	27	75.0	187	1	YK06_AERPE
31	27	75.0	261	1	ZNUB_HAEIN
32	27	75.0	302	1	SC14_CANGA
33	27	75.0	383	1	HKL5_ARATH

ALIGNMENTS

34	27	75.0	474	1	YLEA_ECOLI	P77645	Escherichia
35	27	75.0	748	1	STA4_HUMAN	Q14765	homo sapien
36	27	75.0	749	1	STA4_MOUSE	P42228	mus musculus
37	27	75.0	851	1	STA2_HUMAN	P52630	homo sapien
38	27	75.0	1220	1	C5AC_BACTU	P56955	baclillus th
39	27	75.0	1385	1	C5AA_BACUD	Q45760	baclillus th
40	26	72.2	152	1	RS15_ARCFU	Q29457	archaeoglob
41	26	72.2	160	1	YIBK_HAEIN	P44868	haemophilus
42	26	72.2	199	1	VTAK_LAMB	P03729	bacterioph
43	26	72.2	203	1	CHPE_PSEAE	O87005	pseudomon
44	26	72.2	301	1	SC14_KILUA	P24859	kluyveromy
45	26	72.2	302	1	YDX3_SCHPO	O14166	schizosacch
46	26	72.2	303	1	META_BACHD	Q9AK7	baclillus ha
47	26	72.2	378	1	YGGW_ECOLI	P52062	Escherichia
48	26	72.2	380	1	FD3E_SOYBN	P48625	glycine max
49	26	72.2	400	1	FDH_PSEER	P33160	pseudomon
50	26	72.2	435	1	AM3D_ORRSA	P27933	oryza sativ

RESULT 1
CLD2_MOUSE
ID CLD2_MOUSE STANDARD; PRT; 230 AA.
AC 088552;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-2.
GN CLDN2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC Mammalia; Euteria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98311639; PubMed=9647647; Fujimoto K., Tsukita S.;
RA Furuse M., Fujita K., Hiltaghi T.,
RT "Claudin-1 and -2: novel integral membrane proteins localizing at
tight junctions with no sequence similarity to occludin.";
RL J. Cell Biol. 141:1539-1550(1998).
CC - FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC - SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
EMBL: AF072128; AAC27079.1; -
MGD: MGI:1276110; Cldn2.
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin; 1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN; 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 163 183 POTENTIAL.
SQ SEQUENCE 230 AA: 24483 MW: 38A7C074A1E0D5D2 CRC64;

Query Match 88.9%; Score 32; DB 1; Length 230;
Best Local Similarity 62.5%; Pred. No. 3.3;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

Oy      1 wrxssyxxg 8
        || || |
Db      30 WRTSYSG 37

RESULT  2
EX58_ECOLI      STANDARD;      PRT;      1180 AA.
AC      P08394.
DT      01-AUG-1988 (Rel. 08, Created)
DT      01-AUG-1988 (Rel. 08, Last sequence update)
DT      20-AUG-2001 (Rel. 40, Last annotation update)
DE      EXODEOXYRIBONUCLEASE V BETA CHAIN (EC 3.1.11.5) (EXODEOXYRIBONUCLEASE
DE      V 135 KDA POLYPEPTIDE).
GN      RECB OR RORA OR B2820.
OS      Escherichia coli.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Escherichia.
OX      NCBI_TaxID=562;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      MEDLINE=87066729; PubMed=3537960;
RA      Finch P.W., Storey A., Chapman K.E., Brown K., Hickson I.D.,
RT      Emerson P.T.;
RT      "Complete nucleotide sequence of the Escherichia coli recB gene.";
RN      [2]
RP      Nucleic Acids Res. 14:8573-8582(1986).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=V1000;
RC      MEDLINE=20229837; PubMed=10766864;
RA      Arnold D.A., Kowalczykowski S.C.;
RT      "Facilitated loading of RecA protein is essential to recombination by
RT      RecBCD enzyme.";
RN      J. Biol. Chem. 275:12261-12265(2000).
RN      [4]
RP      SEQUENCE FROM N.A.
RC      STRAIN=K12 / MG1655;
RC      MEDLINE=97426617; PubMed=9278503;
RA      Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA      Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA      Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA      Mau B., Shao Y.;
RT      "The complete genome sequence of Escherichia coli K-12.";
RN      Science 277:1453-1474(1997).
RN      [5]
RP      SEQUENCE OF 1-11 FROM N.A.
RC      MEDLINE=87040734; PubMed=3534791;
RA      Finch P.W., Wilson R.E., Brown K., Hickson I.D., Emerson P.T.;
RT      "Complete nucleotide sequence of the Escherichia coli ptr gene
RT      encoding protease III.";
RN      Nucleic Acids Res. 14:7695-7703(1986).
RN      [6]
RP      SEQUENCE OF 1093-1180 FROM N.A.
RC      MEDLINE=87066730; PubMed=3537961;
RA      Finch P.W., Storey A., Brown K., Hickson I.D., Emerson P.T.;
RT      "Complete nucleotide sequence of recD, the structural gene for the
RT      alpha subunit of Exonuclease V of Escherichia coli.";
RN      Nucleic Acids Res. 14:8583-8594(1986).
RN      [7]
RP      FUNCTION: REQUIRED FOR EFFICIENT DNA REPAIR; IT CATALYZES THE
RN      UNWINDING OF DOUBLE-STRANDED DNA AND THE CLEAVAGE OF SINGLE-
RN      STRANDED DNA AND IT STIMULATES LOCAL GENETIC RECOMBINATION.
RN      .ALL OF THESE ACTIVITIES REQUIRE CONCOMITANT HYDROLYSIS OF ATP.
RN      -1- CATALYTIC ACTIVITY: EXONUCLYTIC CLEAVAGE (IN THE PRESENCE OF
RN      ATP) IN EITHER 5'-TO 3'-OR 3'-TO 5'-DIRECTION TO YIELD 5'-
RN      PHOSPHOLIGONUCLEOTIDES.
RN      -1- SUBUNIT: CONSIST OF THREE SUBUNITS; RECB, RECC AND RECD.
RN      -1- SIMILARITY: BELONGS TO THE UVRD SUBFAMILY OF HELICASES.
RN      -----
RN      This SWISS-PROT entry is copyright. It is produced through a collaboration
RN      between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN      the European Bioinformatics Institute. There are no restrictions on its
RN      use by non-profit institutions as long as its content is in no way
RN      modified and this statement is not removed. Usage by and for commercial

```

```

CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X04581; CAA28250.1; -
DR      EMBL; Af179304; AAD56369.1; -
DR      EMBL; U29581; AAB40467.1; -
DR      EMBL; AE000365; AAC75859.1; -
DR      EMBL; X06227; CAA29577.1; -
DR      EMBL; X04582; CAA28252.1; -
DR      PTR; A25532; NCECX5.
DR      HSSP; P56235; 1PDR.
DR      EcoGene; EG10824; recB.
DR      InterPro; IPR000212; UVRD-helicase.
DR      Pfam; PF00580; UVRD-helicase; 1.
KW      Hydrolase; Nuclease; Exonuclease; Endonuclease; Helicase; ATP-binding;
KW      DNA repair; Complete proteome
FT      NP_BIND      23      30      ATP.
SQ      SEQUENCE      1180 AA; 133958 MW; F9AC331808BF281 CRC64;

Oy      1 wrxssyxxg 8
        || || |
Db      901 WRTSYSG 908

Query Match      88.9%; Score 32; DB 1; Length 1180;
Best Local Similarity      62.5%; Pred No. 16;
Matches      5; Conservative      0; Mismatches      3; Indels      0; Gaps      0;

RESULT  3
CLD1_HUMAN      STANDARD;      PRT;      211 AA.
ID      CLD1_HUMAN
AC      G95832;
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      20-AUG-2001 (Rel. 40, Last annotation update)
DE      CLAUDIN-1 (SENESCENCE-ASSOCIATED EPITHELIAL MEMBRANE PROTEIN).
GN      CLDN1 OR CLD1 OR SEMP1.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      MEDLINE=99132301; PubMed=9931503;
RA      Weissheim K.L., Machl A., Plantzter S., Robertson R., Kublies M.,
RA      Hosier S.;
RT      "SEMP1, a senescence-associated cDNA isolated from human mammary
RT      epithelial cells, is a member of an epithelial membrane protein
RT      superfamily.";
RN      Gene 226:285-295(1999).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      MEDLINE=20290992; PubMed=10828592;
RA      Halford S., Spencer P., Greenwood J., Winton H., Hunt D.M.,
RA      Adamson P.;
RT      "Assignment(1) of claudin-1 (CLDN1) to human chromosome 3q28-->q29
RT      with somatic cell hybrids.";
RN      Cytogenet. Cell Genet. 88:217-217(2000).
RN      [3]
RP      FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
RN      -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
RN      -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
RN      -----
RN      This SWISS-PROT entry is copyright. It is produced through a collaboration
RN      between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN      the European Bioinformatics Institute. There are no restrictions on its
RN      use by non-profit institutions as long as its content is in no way
RN      modified and this statement is not removed. Usage by and for commercial

```

```
CC entitles requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL: AF101051; AADI6433.1; -
DR EMBL: AF115546; AAD22962.1; -
DR EMBL: AF134160; AAF61393.1; -
DR MIM: 603718; -
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin; 1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN.1.
KW Tight Junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 116 136 POTENTIAL.
FT TRANSMEM 164 184 POTENTIAL.
FT CONFLICT 62 62 I -> V (IN REF. 2).
FT CONFLICT 135 135 V -> A (IN REF. 2).
SQ SEQUENCE 211 AA; 22744 MW; 0726900DE6C214F0 CRC64;

Query Match
Best Local Similarity 86.1%; Score 31; DB 1; Length 211;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxxyxg 8
DB 30 WRISYAG 37

RESULT 4
ID YVD3_CAEEL STANDARD: PRT; 551 AA.
AC P53114;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL ZINC METALLOPROTEINASE K04E7.3 PRECURSOR (EC 3.4.24.-).
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Polodeirinae; Caenorhabditis.
OX NCBI_TaxID:6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Nhan M.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12A (ZINC
METALLOPROTEINASE); ALSO KNOWN AS THE ASTACIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL: U39666; AAA80412.1; -
CC HSSP: P00740; IIXA.
DR Wormpep: K04E7.3; CE02798.
DR InterPro: IPR001506; Astacin.
DR InterPro: IPR000859; CUB.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000884; TSP1.
DR InterPro: IPR000130; Zn_MTPeptide.
DR Pfam: PF01400; Astacin; 1.
DR Pfam: PF00431; CUB; 1.
```

```
DR SMART: SM00042; CUB; 1.
DR SMART: SM00209; TSP1; 1.
DR SMART: SM00235; Znmc; 1.
DR PROSITE: PS01180; CUB; FALSE_NEG.
DR PROSITE: PS00022; EGF_1; 1.
DR PROSITE: PS01186; EGF_2; FALSE_NEG.
DR PROSITE: PS50092; TSP1; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
KW Hypothetical protein; Hydrolase; Metalloprotease; Zinc;
KW Signal; EGF-like domain; Glycoprotein.
FT SIGNAL 1 ?
FT CHAIN 1 551
FT DOMAIN 18 28
FT DOMAIN 318 358
FT DOMAIN 359 484
FT DOMAIN 485 535
FT METAL 219 219
FT ACT_SITE 220 220
FT METAL 223 223
FT METAL 229 229
FT METAL 136 136
FT CARBOHYD 378 378
SQ SEQUENCE 551 AA; 61673 MW; 7CFC9A16B56C8B7E CRC64;
```

```
Query Match
Best Local Similarity 86.1%; Score 31; DB 1; Length 551;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 wrxxxyxg 8
DB 376 WRNISYG 383
```

```
RESULT 5
ID VAT_CAMVP STANDARD: PRT; 159 AA.
AC P19818;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE APHID TRANSMISSION PROTEIN (PROTEIN 2).
CN II.
OS Cauliflower mosaic virus (strain PV147) (Camv).
OC Viruses; Retroid viruses; Caulimovirus.
OX NCBI_TaxID:10647;
RN [1]
RP SEQUENCE FROM N.A.
RA Modlanaei N., Volovitch M., Mazzolini L., Yot P.;
RT "Comparison of the predicted secondary structure of aphid
RT transmission factor for transmissible and non-transmissible
RT cauliflower mosaic virus strains.";
RL FEBS Lett. 181:223-228(1985).
CC -1- FUNCTION: THIS PROTEIN IS INVOLVED IN VIRUS TRANSMISSION.
CC -1- SIMILARITY: BELONGS TO THE CAULIMOVIRUSES ORF II FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL: M37581; AAA96697.1; -
DR SEQUENCE 159 AA; 17843 MW; 196DF6D1F9D0B15A CRC64;
```

```
Query Match
Best Local Similarity 83.3%; Score 30; DB 1; Length 159;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```

QY      1 wrxsyxyg 8
DB      61 WRINSYFG 68

RESULT  6
ZNUB_ECOLI STANDARD; PRT; 261 AA.
AC P39832: P76286;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HIGH-AFFINITY ZINC UPTAKE SYSTEM MEMBRANE PROTEIN ZNUB.
GN ZNUB OR B1859.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitakawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Motoki T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
RA Saito N., Sampei G., Seki Y., Sivasubramanian S., Tagami H.,
RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horinouchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map."
RL DNA Res. 3:379-392(1996).
RN [3]
RP SEQUENCE OF 1-92 FROM N.A.
RC STRAIN=K12 / EMG2;
RA Robinson K., O'Keefe T., Church G.M.;
RT Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 80-261 FROM N.A.
RC MEDLINE=88314937; PubMed=2842314;
RA Shigaawa H., Makino K., Amemura M., Kimura S., Iwasaki H., Nakata A.;
RT "Structure and regulation of the Escherichia coli ruv operon involved
RT in DNA repair and recombination."
RL J. Bacteriol. 170:4322-4329(1988).
RN [5]
RP IDENTIFICATION.
RC MEDLINE=95075659; PubMed=7984428;
RA Borodovsky M., Rudd K.E., Koonin E.V.;
RT "Intrinsic and extrinsic approaches for detecting genes in a
RT bacterial genome."
RL Nucleic Acids Res. 22:4756-4767(1994).
RN [6]
RP CHARACTERIZATION.
RC MEDLINE=98343803; PubMed=9680209;
RA Patzer S.I., Hantke K.;
RT "The znuABC high-affinity zinc uptake system and its regulator zur in
RT Escherichia coli."
RL Mol. Microbiol. 28:1199-1210(1998).
CC -1- FUNCTION: INVOLVED IN THE HIGH-AFFINITY ZINC UPTAKE TRANSPORT
CC SYSTEM.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE ABC-3 SUBFAMILY OF INTEGRAL MEMBRANE
CC PROTEINS. STRONG, TO H. INFLUENZAE ZNUB.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF000280; AAC74929.1; -.
DR EMBL; D90828; BAA15667.1; -.
DR EMBL; D90829; BAA15670.1; -.
DR EMBL; U38702; AAA81031.1; -.
DR EMBL; M21298; -. NOT_ANNOTATED_CDS.
DR EcoGene; EG12368; znuB.
DR InterPro; IPR001626; ABC-3.
DR InterPro; IPR001872; Lipo_sig_Pase.
DR Pfam; PF00950; ABC-3; 1.
DR PRINTS; PR00781; LIPOSIGPTASE.
KW Transport; Zinc; Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 54 74 POTENTIAL.
FT TRANSMEM 84 104 POTENTIAL.
FT TRANSMEM 122 142 POTENTIAL.
FT TRANSMEM 178 198 POTENTIAL.
FT TRANSMEM 214 234 POTENTIAL.
FT TRANSMEM 236 256 POTENTIAL.
SQ SEQUENCE 261 AA; 27728 MW; 78382B2EACCE1490 CRC64;

Query Match 83.3%; Score 30; DB 1; Length 261;
Best Local Similarity 62.5%; Pred. No. 10;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 wrxsyxyg 8
DB      28 WRRMSYFG 35

RESULT  7
CLD2_HUMAN STANDARD; PRT; 230 AA.
ID CLD2_HUMAN
AC P57339;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE CLAUDIN-2.
GN CLDN2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Epithelium;
RA Reinacker H.-C.; Sakaguchi T.; Golden H.M.;
RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,
RA Qin W.X., Huang Y., Qin X.K., Qian L.F., He L.P., Li H.N., Yu Y.,
RA Yu J., Han L.H.;
RT "Novel human cDNA clone with function of inhibiting cancer cell
RT growth."
RT Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```



```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF250558; AAF98151.1; -.
DR EMBL: AF177340; AAG17984.1; -.
DR InterPro: IPR001832; CLAUDIn.
DR Pfam: PF00822; PM22.Claudin.1.
DR PROSITE: PS01346; CLAUDIN.1.
KW Tight Junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 163 183 POTENTIAL.
SQ SEQUENCE 230 AA; 24548 MW; 52CA642DA62B700 CRC64;

Query Match 80.6%; Score 29; DB 1; Length 230;
Best Local Similarity 50.0%; Pred. No. 15;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxsyxg 8
   |||
Db 30 WRTSSYVG 37

RESULT 8
AOPH_MYCTU STANDARD; PRT: 454 AA.
ID AOPH_MYCTU
DT 15-DEC-1998 (Rel. 37, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE FLAVIN-CONTAINING MONOAMINE OXIDASE RV3170 (EC 1.4.3.-).
GN RV3170 OR MT3259 OR MTV014.14.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_Taxid:1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RC MEDLINE:98295987; PubMed:9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Baudock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jurgels K., Kiroh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Ruter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultun J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolenko M.D., Salzberg S.L.,
RA Delcher A., Uitterlinden T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai M.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
CC -1- COFACTOR: FAD (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE FLAVIN MONOAMINE OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE006381; AAK05600.1; -.
DR PROSITE: PS00187; TPP_ENZYMES; FALSE_NEG.
KW Lyase; Flavoprotein; Thiamine pyrophosphate; Complete proteome.
SQ SEQUENCE 822 AA; 93363 MW; C686D569D38E22F CRC64;
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AL021646; CAA16635.1; ALT_INIT.
DR EMBL: AE007139; AAK47598.1; -.
DR TIGR: MT3259; -.
DR TuberculList: RV3170; -.
DR InterPro: IPR002937; Amino-oxidase.
DR InterPro: IPR000205; NAD-binding.
DR Pfam: PF01593; Amino-oxidase.1.
KW Hypothetical protein; oxidoreductase; Flavoprotein; FAD;
KW Complete proteome.
FT NP_BIND 14 69 FAD (ADP PART) (POTENTIAL).
FT NP_BIND 14 69 FAD (ADP PART) (POTENTIAL).
SQ SEQUENCE 454 AA; 49136 MW; 6C1AE9B97FB2F435F CRC64;

Query Match 80.6%; Score 29; DB 1; Length 454;
Best Local Similarity 50.0%; Pred. No. 29;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 wrxxsyxg 8
   |||
Db 312 WRASGYSG 319

RESULT 9
XFP_LACLA STANDARD; PRT: 822 AA.
ID XFP_LACLA
AC Q9CFH4;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE XYLULOSE-5-PHOSPHATE/FRUCTOSE-6-PHOSPHATE PHOSPHOKETOYLASE
DE (EC 4.1.2.9) (EC 4.1.2.22).
GN XFP OR PTK.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_Taxid:1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RC MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Winkler P., Manger S., Jallion O., Malarre K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403.";
RL Genome Res. 11:731-753(2001).
CC -1- FUNCTION: PHOSPHOKETOYLASE USING BOTH FRUCTOSE 6-PHOSPHATE AND
CC XYLULOSE 5-PHOSPHATE AS SUBSTRATE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: D-XYLULOSE 5-PHOSPHATE + PHOSPHATE = ACETYL
CC -1- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + PHOSPHATE = ACETYL
CC -1- CATALYTIC ACTIVITY: D-FRUCTOSE 6-PHOSPHATE + PHOSPHATE = ACETYL
CC -1- CATALYTIC ACTIVITY: D-ERYTHROSE 4-PHOSPHATE + H(2)O.
CC -1- COFACTOR: THIAMINE PYROPHOSPHATE (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE XFP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE006381; AAK05600.1; -.
DR PROSITE: PS00187; TPP_ENZYMES; FALSE_NEG.
KW Lyase; Flavoprotein; Thiamine pyrophosphate; Complete proteome.
SQ SEQUENCE 822 AA; 93363 MW; C686D569D38E22F CRC64;
```

```

Best Local Similarity 50.0%; Pred. No. 51;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsyxg 8
   11:11
Db 17 WRATYLG 24

RESULT 10
LHA2_ECTHL STANDARD; PRT; 65 AA.
ID LHA2_ECTHL
AC P80103;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, last sequence update)
DT 01-FEB-1994 (Rel. 28, last annotation update)
DE LIGHT-HARVESTING PROTEIN B800/830/1020, ALPHA-2 CHAIN (EMS-ALPHA-2)
DE (ANTENNA PIGMENT PROTEIN, ALPHA-2 CHAIN).
OS Ectochlorodospira halophilis.
OC Bacteria; Proteobacteria; gamma subdivision; Ectochlorodospiraceae;
OC Halorhodospira.
OX NCBI_TaxID=1052;
RN [1]
RP SEQUENCE.
RC STRAIN=DSM 1059;
RX MEDLINE=92249336; PubMed=15777009;
RA Wagner-Huber R., Brunisholz R.A., Bissig I., Frank G., Suter F.,
RA Zuber H.;
RT "The primary structure of the antenna polypeptides of
RT Ectochlorodospira halophilis and Ectochlorodospira halophila. Four
RT core-type antenna polypeptides in E. halophilis and E. halophila."
RL Eur. J. Biochem. 205:917-925(1992).
CC -I- FUNCTION: ANTENNA COMPLEXES ARE LIGHT-HARVESTING SYSTEMS, WHICH
CC TRANSFER THE EXCITATION ENERGY TO THE REACTION CENTERS.
CC -I- SUBUNIT: THE CORE COMPLEX IS FORMED BY DIFFERENT ALPHA AND BETA
CC CHAINS, BINDING BACTERIOCHLOROPHYLL MOLECULES, AND ARRANGED
CC MOST PROBABLY IN TETRAMERIC STRUCTURES DISPOSED AROUND THE
CC REACTION CENTER. THE NON-PIGMENTED GAMMA CHAINS MAY CONSTITUTE
CC ADDITIONAL COMPONENTS.
CC -I- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. INNER MEMBRANE.
CC PIR: S23164; S23164.
DR InterPro: IPR002361; Antenna_comp_alpha.
DR InterPro: IPR000066; LHC.
DR Pfam: PF00556; LHC. 1.
DR PROSITE: PS00968; ANTENNA_COMP_ALPHA; 1.
KW Antenna complex; Light-harvesting polypeptide; Transmembrane;
KW Magnesium; Bacteriochlorophyll; Inner membrane.
FT DOMAIN 1 13 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 14 34 PERIPLASMIC (POTENTIAL).
FT DOMAIN 35 65 AXIAL LIGAND TO THE BACTERIOCHLOROPHYLL
FT METAL 29 29 MAGNESIUM (POTENTIAL).
SQ SEQUENCE 65 AA; 7688 MW; 55A4C306748BD9A CRC64;

Query Match 77.8%; Score 28; DB 1; Length 65;
Best Local Similarity 50.0%; Pred. No. 7.5;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsyxg 8
   11:11
Db 58 WRSTYDG 65

RESULT 11
OOX2_SULAC STANDARD; PRT; 168 AA.
ID OOX2_SULAC
AC P39479;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, last sequence update)
DT 15-JUL-1998 (Rel. 36, last annotation update)
DE OXINOL OXIDASE POLYPEPTIDE II (EC 1.9.3.-) (CYTOCHROME AA3 SUBUNIT 2)
DE (OXIDASE AA(3) SUBUNIT 2).
SOXAX.
```

```

OS Sulfolobus acidocaldarius.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=2285;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33909 / NCIB 11770 / DSM 639;
RX MEDLINE=92192013; PubMed=1372250;
RA Luebben M., Kolmerer B., Saraste M.;
RT "An archaebacterial terminal oxidase combines core structures of two
RT mitochondrial respiratory complexes."
RL EMBO J. 11:805-812(1992).
CC -I- FUNCTION: THE TERMINAL OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX.
CC -I- FUNCTION: SUBUNIT 2 TRANSFERS THE ELECTRONS FROM CALDIIRIELLA
CC OXINOL TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1 THAT IS
CC FORMED BY HEME A3 AND CU(B).
CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -I- SIMILARITY: SOME, TO MITOCHONDRIAL OR BACTERIAL COX2 SUBUNITS.
CC BUT LACK HEME-BINDING DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X62643; CAA44509.1; -.
DR PIR; S21041; S21041.
KW oxidoreductase; Transmembrane; Respiratory chain; Electron transport.
FT TRANSMEM 9 31 POTENTIAL.
SQ SEQUENCE 168 AA; 18874 MW; 7B5C98CA2AD06846 CRC64;

Query Match 77.8%; Score 28; DB 1; Length 168;
Best Local Similarity 50.0%; Pred. No. 19;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxxsyxg 8
   11:11
Db 147 WRDAEYAG 154

RESULT 12
CLD7_RAT STANDARD; PRT; 191 AA.
ID CLD7_RAT
AC Q9Z1L1;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, last sequence update)
DT 30-MAY-2000 (Rel. 39, last annotation update)
DE CLAUDIN-7 (FRAGMENT).
DE CLDN7.
GN CLDN7.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Keen T.J., Inglehearn C.F.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -I- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```

CC -----
DR EMBL: AJ011811: CAA09790.1; -.
DR InterPro: IPR001832: Claudin.
DR InterPro: IPR000729: PMP22_Claudin.
DR Pfam: PF00882: PMP22_Claudin; 1.
DR PRINTS: PR01077: CLAUDIN.
DR PROSITE: PS01346: CLAUDIN; 1.
KW Tight junction; Transmembrane.
FT NON_TER 1 1
FT TRANSMEM 62 82 POTENTIAL.
FT TRANSMEM 100 120 POTENTIAL.
FT TRANSMEM 141 161 POTENTIAL.
FT TRANSMEM 164 184 POTENTIAL.
SO SEQUENCE 191 AA: 20366 MW: 838445908DFFFA1A CRC64;

Query Match      77.8%; Score 28; DB 1; Length 191;
Best Local Similarity 50.0%; Pred. No. 21;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 WXXXXXXG 8
DB 10 WQSSYAG 17

RESULT 13
CLD1_MOUSE STANDARD; PRT; 211 AA.
AC O88551;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-1.
CLDN1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID:10090;
CX [1]
RN SEQUENCE FROM N.A.
RP MEDLINE:96311639; PubMed:9647647;
RX Furuse M., Fujita K., Hilder T., Fujimoto K., Tsukita S.;
RT "Claudin-1 and -2: novel integral membrane proteins localizing at
tight junctions with no sequence similarity to occludin.";
RL J. Cell Biol. 141:1539-1550(1998).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: AF072127: AAC27078.1; -.
DR MGD: MGI:1276109: Clcn1.
DR InterPro: IPR001832: Claudin.
DR InterPro: IPR000729: PMP22_Claudin.
DR Pfam: PF00882: PMP22_Claudin; 1.
DR PRINTS: PR01077: CLAUDIN.
DR PROSITE: PS01346: CLAUDIN; 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 116 136 POTENTIAL.
FT TRANSMEM 164 184 POTENTIAL.
SO SEQUENCE 211 AA: 22881 MW: BEF896FA62DBB6F0 CRC64;

Query Match      77.8%; Score 28; DB 1; Length 211;
Best Local Similarity 50.0%; Pred. No. 23;

```

```

Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 WXXXXXXG 8
DB 30 WKIYSYAG 37

RESULT 14
CLD1_RAT STANDARD; PRT; 211 AA.
ID CLD1_RAT
AC P56745;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-1.
GN CLDN1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID:10116;
CX [1]
RN SEQUENCE FROM N.A.
RP STRAIN-SPRAGUE-DAWLEY;
RA Gregory M., Dufresne J., Cyr D.G.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: AF195500: AAF04850.1; -.
DR InterPro: IPR001832: Claudin.
DR InterPro: IPR000729: PMP22_Claudin.
DR Pfam: PF00882: PMP22_Claudin; 1.
DR PRINTS: PR01077: CLAUDIN.
DR PROSITE: PS01346: CLAUDIN; 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 116 136 POTENTIAL.
FT TRANSMEM 164 184 POTENTIAL.
SO SEQUENCE 211 AA: 22871 MW: 3595931636631DDC CRC64;

Query Match      77.8%; Score 28; DB 1; Length 211;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 WXXXXXXG 8
DB 30 WKIYSYAG 37

RESULT 15
CLD7_HUMAN STANDARD; PRT; 211 AA.
ID CLD7_HUMAN
AC G95471;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-7.
GN CLDN7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID:9606;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Colon adenocarcinoma;
RC Keen T.J.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ011497; CAA09626.1; -
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin; 1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN; 1.
DR TIGHT junction: Transmembrane.
KW TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 118 138 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
SQ SEQUENCE 211 AA; 22390 MW; 7F3CC1B963D912E1 CRC64;

Query Match 77.8%; Score 28; DB 1; Length 211;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxsxyxg 8
   1: 111
Db 30 WOMSSYAG 37

RESULT 16
CLUD_MOUSE STANDARD; PRT; 211 AA.
ID CLUD_MOUSE
AC O9261;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CLAUDIN-7.
GN CLDN7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99110921; PubMed=9892664;
RA Morita K., Furuse M., Fujimoto K., Tsukita S.;
RT "Claudin multigene family encoding four-transmembrane domain protein
RT components of tight junction strands."
RL Proc. Natl. Acad. Sci. U.S.A. 96:511-516(1999).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF087825; AAD09760.1; -

```

```

DR MGD: MGI:1859285; Clud7.
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin; 1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN; 1.
DR TIGHT junction: Transmembrane.
KW TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 120 140 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
SQ SEQUENCE 211 AA; 22359 MW; 4FE87FE3A57AC9F29 CRC64;

Query Match 77.8%; Score 28; DB 1; Length 211;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxsxyxg 8
   1: 111
Db 30 WOMSSYAG 37

RESULT 17
SPS2_MOUSE STANDARD; PRT; 452 AA.
ID SPS2_MOUSE
AC P97364;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SELENIDE, WATER DIKINASE 2 (EC 2.7.9.3) (SELENOPHOSPHATE SYNTHETASE 2)
DE (SELENIUM DONOR PROTEIN 2).
GN SPS2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96017645; PubMed=7588067;
RA Guimaraes M.J., Bazan J.F., Zlotnick A., Wiles M.V., Grimaldi J.C.,
RA Lee F., McClanahan T.;
RT "A new approach to the study of haematopoietic development in the
RT yolk sac and embryoid bodies."
RL Development 121:3335-3346(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97140286; PubMed=8986768;
RA Guimaraes M.J., Peterson D., Vicari A., Cocks B.G., Copeland N.G.,
RA Gilbert D.J., Jenkins N.A., Ferrick D.A., Kastelein R., Bazan J.F.,
RA Zlotnick A.;
RT "Identification of a novel seid homolog from eukaryotes, bacteria,
RT and archaea: is there an autoregulatory mechanism in selenocysteine
RT metabolism?";
RL Proc. Natl. Acad. Sci. U.S.A. 93:15086-15091(1996).
CC -1- FUNCTION: SYNTHESIZES SELENOPHOSPHATE FROM SELENIDE AND ATP.
CC -1- CATALYTIC ACTIVITY: ATP + SELENIDE + H(2)O = AMP + SELENOPHOSPHATE
CC + PHOSPHATE.
CC -1- COFACTOR: SELENOCYSTEINE. THE ACTIVE-SITE SELENOCYSTEINE IS
CC ENCODED BY THE OPAL CODON, UGA.
CC -1- SIMILARITY: BELONGS TO THE SELENOPHOSPHATE SYNTHETASE 1 FAMILY.
CC CLASS I SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U43285; AAC53024.1; -
DR MGD: MGI:108388; Sps2.

```

```

DR InFePro: IPR000728; AIRS_related.
KW Transferase: Selenium; Selenocysteine; ATP-binding.
FT ACT_SITE 63 63 POTENTIAL.
FT SE_CYS 63 63
FT SITE 66 66 IMPORTANT FOR CATALYTIC ACTIVITY (BY
SIMILARITY).
FT NP_BIND 322 328 ATP (POTENTIAL).
FT DOMAIN 2 9 POLY-ALA.
FT DOMAIN 433 440 POLY-ALA.
SQ SEQUENCE 452 AA; 47786 MW; 9DA6F7250CFE80E4 CRC64;

Query Match 77.8%; Score 28; DB 1; Length 452;
Best Local Similarity 50.0%; Pred. No. 48;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsyxg 8
DB 52 WRLTSFSG 59

RESULT 18
AOL1_THEAO STANDARD; PRT; 513 AA.
AC P08554;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE AQUALYSIN I PRECURSOR (EC 3.4.21.-).
GN PSTI.
OS Thermus aquaticus.
OC Bacteria; Thermus/Delnoococcus group; Thermus group; Thermus.
OX NCBI_TaxID:271;
RN (1)
RN SEQUENCE FROM N.A., AND SEQUENCE OF 15-23.
RC STRAIN-YT1;
RX MEDLINE=90216674; PubMed=2182621;
RA Terada I., Kwon S.-T., Miyata Y., Matsuzawa H., Ohta T.;
RT "Unique precursor structure of an extracellular protease, aqualysin
I, with NH2- and COOH-terminal pro-sequences and its processing in
Escherichia coli.";
RL J. Biol. Chem. 265:6576-6581(1990).
RN (2)
RN SEQUENCE OF 75-442 FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-YT1;
RX MEDLINE=88225062; PubMed=3286255;
RA Kwon S.-T., Terada I., Matsuzawa H., Ohta T.;
RT "Nucleotide sequence of the gene for aqualysin I (a thermophilic
alkaline serine protease) of Thermus aquaticus YT-1 and
characteristics of the deduced primary structure of the enzyme.";
RL Eur. J. Biochem. 173:491-497(1988).
RN (3)
RN SEQUENCE OF 128-170.
RX MEDLINE=88151937; PubMed=3162211;
RA Matsuzawa H., Tokugawa K., Hamocki M., Mizoguchi M., Taguchi H.,
Terada I., Kwon S.-T., Ohta T.;
RT "Purification and characterization of aqualysin I (a thermophilic
alkaline serine protease) produced by Thermus aquaticus YT-1.";
RL Eur. J. Biochem. 171:441-447(1988).
RN (4)
RN FUNCTION: AQUALYSIN I IS A THERMOPHILIC ALKALINE SERINE PROTEASE.
RN THE OPTIMAL TEMPERATURE FOR ITS CASEINOLYTIC ACTIVITY IS 80
DEGREES CELSIUS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DEVELOPMENTAL STAGE: SECRETED FROM THE EARLY STATIONARY PHASE
UNTIL THE TIME THE CELLS CEASE TO GROW.
CC -1- PTM: THE N- AND C-TERMINAL PRO-SEQUENCES ARE REMOVED THROUGH THE
PROTEOLYTIC ACTIVITY OF AQUALYSIN I ITSELF, IN THAT ORDER. THE
C-TERMINAL PRO-SEQUENCE IS REQUIRED FOR TRANSLOCATION OF THE
PROTEASES ACROSS THE OUTER MEMBRANE.
CC -1- PTM: TWO DISULFIDE BONDS ARE PRESENT.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
SUBTILASE FAMILY.

```

```

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D90108; BAAL1435.1;
CC EMBL: X07734; CAA30559.1;
CC PIR: S00620; S00620.
CC PTR: A35742; A35742.
CC HSSP: P06873; 3PRK.
CC MEROPS: S08.051; -.
CC InFePro: IPR000209; Peptidase-S8.
CC Pfam: PF00082; Peptidase_S8; 1.
CC PRINTS: PR00723; SUBTILISIN.
CC PROSITE: PS00136; SUBTILASE_ASP; 1.
CC PROSITE: PS00137; SUBTILASE_HIS; 1.
CC PROSITE: PS00138; SUBTILASE_SER; 1.
CC K01 Hydrolyase; Serine protease; zymogen; signal.
FT SIGNAL 1 14
FT PROPEP 15 127
FT CHAIN 128 408 AQUALYSIN I.
FT PROPEP 409 513
FT ACT_SITE 166 166 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 197 197 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 349 349 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 513 AA; 53913 MW; DDFDFE6DA50B785 CRC64;

Query Match 77.8%; Score 28; DB 1; Length 513;
Best Local Similarity 50.0%; Pred. No. 54;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsyxg 8
DB 495 WRIVAYSG 502

RESULT 19
YNA8_YEAST STANDARD; PRT; 669 AA.
ID YNA8_YEAST
AC P53983;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOPHETICAL 76.7 KDA PROTEIN IN SPO1-SIS1 INTERGENIC REGION.
GN YNL008C OR N2874.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID:4932;
RN (1)
RN SEQUENCE OF 1-102 FROM N.A.
RA Andre B., Iraqui Housaini I., Urrestarazu L.A., Vissers S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN (2)
RN SEQUENCE OF 371-669 FROM N.A.
RA Dolignon F., Crouzet M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z71284; CAA95868.1;

```

```

DR EMBL: Z71283; CA95867.1; -
DR SCD; S0004953; INL008C.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 72 92 POTENTIAL.
FT TRANSMEM 120 140 POTENTIAL.
FT TRANSMEM 150 170 POTENTIAL.
FT TRANSMEM 175 195 POTENTIAL.
FT TRANSMEM 209 229 POTENTIAL.
FT TRANSMEM 234 254 POTENTIAL.
FT TRANSMEM 271 291 POTENTIAL.
SQ SEQUENCE 669 AA; 76741 MW; 1BB4DCAC74E8FF63 CRC64;

Query Match 77.8%; Score 28; DB 1; Length 669;
Best Local Similarity 50.0%; Pred. No. 69;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsyxg 8
| | : | |
Db 310 WRSNNYNG 317

RESULT 20
HMDL PICJA STANDARD; PRT; 934 AA.
AC 074164;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE 3-HYDROXY-3-METHYLGUTARYL-COENZYME A REDUCTASE (EC 1.1.1.34) (HMG-COA
DE REDUCTASE).
GN HMG.
OS Pichia jadinii (Yeast) (Candida utilis).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OX NCBI_TaxID=4903;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-98320604; PubMed-9647847;
RA Shimada H., Kondo K., Fraser P.D., Miura Y., Saito T., Misawa N.;
RT "Increased carotenoid production by the food yeast Candida utilis
RT through metabolic engineering of the isoprenoid pathway.";
RL Appl. Environ. Microbiol. 64:2676-2680(1998).
CC -!- FUNCTION: INVOLVED IN THE CONTROL OF CHOLESTEROL BIOSYNTHESIS. IT
CC IS THE RATE-LIMITING ENZYME OF THE STEROL BIOSYNTHESIS.
CC -!- CATALYTIC ACTIVITY: (R)-MEVALONATE + COA + 2 NADP(+) = (S)-3-
CC HYDROXY-3-METHYLGUTARYL-COA + 2 NADPH.
CC -!- PATHWAY: CHOLESTEROL BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
CC RETICULUM.
CC -!- SIMILARITY: BELONGS TO THE HMG-COA REDUCTASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB012603; BAA31937.1; -
DR InterPro: IPR002302; HMG-CoA_red.
DR InterPro: IPR000731; HMGCR_patched_5TM.
DR Pfam: PF00368; HMG-CoA_red; 1.
DR PRINTS: PR00071; HMGCOARDTASE.
DR PROSITE: PS00066; HMG_COA_REDUCTASE_1; 1.
DR PROSITE: PS00318; HMG_COA_REDUCTASE_2; 1.
DR PROSITE: PS01192; HMG_COA_REDUCTASE_3; 1.
DR PROSITE: PS50065; HMG_COA_REDUCTASE_4; 1.
DR PROSITE: PS50156; SSD; 1.
DR Oxidoreductase; Glycoprotein; Endoplasmic reticulum; Transmembrane;
KW Cholesterol biosynthesis; NADP.
FT DOMAIN 1 442 MEMBRANE-BOUND.

```

```

FT DOMAIN 443 521 LINKER.
FT DOMAIN 522 934 CATALYTIC.
FT TRANSMEM 112 132 POTENTIAL.
FT TRANSMEM 142 162 POTENTIAL.
FT TRANSMEM 257 277 POTENTIAL.
FT TRANSMEM 335 355 POTENTIAL.
FT TRANSMEM 422 442 POTENTIAL.
FT ACT_SITE 618 618 BY SIMILARITY.
FT ACT_SITE 828 828 BY SIMILARITY.
FT ACT_SITE 924 924 GENERAL BASE (BY SIMILARITY).
FT CARBOHYD 361 361 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 364 364 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 682 682 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 934 AA; 101233 MW; 3C3E43FC5623601C CRC64;

Query Match 77.8%; Score 28; DB 1; Length 934;
Best Local Similarity 50.0%; Pred. No. 95;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsyxg 8
| | : | |
Db 79 WRSRAVHG 86

```

Search completed: January 14, 2002, 07:40:39
Job time: 505 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:39:44 ; Search time 112.89 Seconds
(without alignments)
10.366 Million cell updates/sec

Title: 09-185908-1E
Perfect score: 36
Sequence: 1 wrxsyxxg 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database :
1: SPREMBL_17:*
2: SP_archaea:*
3: SP_bacteria:*
4: SP_fungi:*
5: SP_human:*
6: SP_invertebrate:*
7: SP_mammal:*
8: SP_mmc:*
9: SP_organelle:*
10: SP_plant:*
11: SP_prodent:*
12: SP_virus:*
13: SP_vertebrate:*
14: SP_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	88.9	65	2 Q9RDQ2	Q9rdq2 streptomyc
2	32	88.9	65	2 Q9RD18	Q9rd18 streptomyc
3	32	88.9	65	2 Q9RD17	Q9rd17 streptomyc
4	32	88.9	1208	2 Q9KPP6	Q9kpp6 vibrio chol
5	31	86.1	278	1 Q58458	Q58458 pyrococcus
6	30	83.3	274	10 Q9SW79	Q9sw79 triticum ae
7	30	83.3	613	2 Q9A956	Q9a956 caulobacter
8	29	80.6	193	11 Q9ET38	Q9et38 mus musculu
9	29	80.6	217	5 Q97402	Q97402 phaedin coc
10	29	80.6	479	12 Q9QON6	Q9qon6 sugarcane y
11	29	80.6	479	12 Q9JH75	Q9jht5 sugarcane y
12	29	80.6	497	4 Q9BRK6	Q9brt6 homo sapien
13	29	80.6	519	2 Q9RKBS	Q9rks5 streptomyc
14	29	80.6	648	10 Q49559	Q49559 arabidopsis
15	29	80.6	822	2 Q9CFH4	Q9cfh4 lactococcus
16	29	80.6	1083	2 Q9RPH6	Q9rph6 mycobacteri
17	28	77.8	157	4 Q95978	Q95978 homo sapien
18	28	77.8	197	10 Q9FHV6	Q9fhv6 arabidopsis
19	28	77.8	206	2 Q9RRD0	Q9rrd0 delinococcus

20	28	77.8	211	4 Q9BYN0	Q9byn0 homo sapien
21	28	77.8	213	5 Q9NAQ9	Q9naq9 caenorhabdi
22	28	77.8	272	10 Q9LTV2	Q9ltv2 arabidopsis
23	28	77.8	366	2 Q9AAC7	Q9aac7 caulobacter
24	28	77.8	432	5 Q9NFP2	Q9nfp2 plasmodium
25	28	77.8	700	10 Q9T0J6	Q9t0j6 arabidopsis
26	28	77.8	988	2 Q83345	Q83345 treponema p
27	28	77.8	7107	5 Q9VAF7	Q9vaf7 dtrosophila
28	27	75.0	64	12 Q83157	Q83157 caullflower
29	27	75.0	68	2 Q9KRD6	Q9krd6 streptomyc
30	27	75.0	99	12 Q83162	Q83162 caullflower
31	27	75.0	115	12 Q64840	Q64840 human adeno
32	27	75.0	154	2 Q9K199	Q9k199 netisseria m
33	27	75.0	154	2 Q9JSM8	Q9jsm8 netisseria m
34	27	75.0	158	2 Q9CCU7	Q9ccu7 mycobacteri
35	27	75.0	159	12 Q83179	Q83179 caullflower
36	27	75.0	159	12 Q66159	Q66159 caullflower
37	27	75.0	159	12 Q83166	Q83166 caullflower
38	27	75.0	159	12 Q9W133	Q9w133 caullflower
39	27	75.0	169	2 Q07698	Q07698 mycobacteri
40	27	75.0	175	8 Q79714	Q79714 rhea ametic
41	27	75.0	199	10 Q9FVVS	Q9fvvs arabidopsis
42	27	75.0	241	2 Q913C4	Q913c4 pseudomonas
43	27	75.0	246	5 Q44142	Q44142 caenorhabdi
44	27	75.0	260	2 Q9KQB7	Q9kqb7 vibrio chol
45	27	75.0	261	2 Q9CP25	Q9cp25 pasteurella
46	27	75.0	262	2 Q9HT72	Q9ht72 pseudomonas
47	27	75.0	266	10 Q9AMX0	Q9amx0 oryza sativ
48	27	75.0	270	5 Q45906	Q45906 caenorhabdi
49	27	75.0	303	10 Q64761	Q64761 arabidopsis
50	27	75.0	323	2 Q9HTT5	Q9htt5 pseudomonas

ALIGNMENTS

RESULT 1
ID Q9RDQ2 PRELIMINARY; PRT; 65 AA.
AC Q9RDQ2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE HYPOTHETICAL 6.9 KDA PROTEIN.
GN SCA47.10.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteriia; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Kienast H., Hopwood D.A.;
RA Kienast H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
KW EMBL: AL133423; CAB62714.1; -;
SO Hypothetical protein.
SEQUENCE 65 AA; 6944 MW; F283FAL5A0650DCE CRC64;

```

Query Match      88.9%; Score 32; DB 2; Length 65;
Best Local Similarity 62.5%; Pred. No. 4.8;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxsyxg 8
   || || |
Db 12 WRRSSYSG 19

RESULT 2
O9RD18 PRELIMINARY; PRT; 65 AA.
AC O9RD18;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE HYPOTHETICAL 7.1 KDA PROTEIN.
GN SCC57A.09C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmid and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL136519; CAB66277.1; -.
KW Hypothetical protein.
SQ SEQUENCE 65 AA; 6959 MW; 1F74C265B9572610 CRC64;

Query Match      88.9%; Score 32; DB 2; Length 65;
Best Local Similarity 62.5%; Pred. No. 4.8;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxsyxg 8
   || || |
Db 7 WRRSSYSG 14

RESULT 3
O9RD17 PRELIMINARY; PRT; 65 AA.
AC O9RD17;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE HYPOTHETICAL 7.1 KDA PROTEIN.
GN SCC57A.10C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.

```

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmid and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL136519; CAB66278.1; -.
KW Hypothetical protein.
SQ SEQUENCE 65 AA; 7127 MW; 4C9A4476C44B727A CRC64;

Query Match      88.9%; Score 32; DB 2; Length 65;
Best Local Similarity 62.5%; Pred. No. 4.8;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxsyxg 8
   || || |
Db 7 WRRSSYSG 14

RESULT 4
O9KPP6 PRELIMINARY; PRT; 1208 AA.
ID O9KPP6;
AC O9KPP6;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE EXODEOXYRIBONUCLEASE V, 135 KDA SUBUNIT.
GN VC2320.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
DR EMBL; AE004302; AAF95464.1; -.
DR TIGR; VC2320; -.
DR InterPro; IPR000212; UvrD-helicase.
DR Pfam; PF00580; UvrD-helicase; 2.
KW Complete proteome.
SQ SEQUENCE 1208 AA; 135860 MW; F3C3BF891435C18D CRC64;

Query Match      88.9%; Score 32; DB 2; Length 1208;
Best Local Similarity 62.5%; Pred. No. 97;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxsyxg 8
   || || |
Db 906 WRVTSYSG 913

RESULT 5
O58458

```

```
ID 058458 PRELIMINARY: PRT: 278 AA.
AC 058458;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DE 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE HYPOTHEtical 31.0 KDA PROTEIN PH0727.
GN PH0727.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
NCBI_TaxID=53953;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
MEDLINE=98344137; PubMed=9679194;
RA Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shikuya H., Kikuchi H.;
RT Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
RL DVA Res. 5:55-76(1998).
DR EMBL: AP000003; BAA29818.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 278 AA; 31004 MW; EE783D23A4E28FA1 CRC64;

Query Match 86.1%; Score 31; DB 1; Length 278;
Best Local Similarity 62.5%; Pred. No. 35;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 WXXXXXX 8
DB 99 WRVSSYMG 106

RESULT 6
Q9SW79 PRELIMINARY: PRT: 274 AA.
AC Q9SW79;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ALTERNATIVE OXIDASE (FRAGMENT).
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticeae; Triticum.
NCBI_TaxID=4565;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=MACHETE; TISSUE=CHLORAMPHENICOL-TREATED SEEDLINGS;
RA Zhang Q., Mischis L., Wiskich J.T.;
RT "Respiratory Responses of pea and wheat seedlings to chloramphenicol
RT treatment."
RN Aust. J. Plant Physiol. 23:583-592(1996).
RN 121
RP SEQUENCE FROM N.A.
RC STRAIN=MACHETE; TISSUE=CHLORAMPHENICOL-TREATED SEEDLINGS;
RA Zhang Q., Wiskich J.T.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF174004; AAD51707.1; -
DR InterPro: IPR002680; AOX.
DR Pfam: PF01786; AOX. 1.
FT NON_TER 1
SQ SEQUENCE 274 AA; 31794 MW; 2271BAED9E0D6B7C CRC64;

Query Match 83.3%; Score 30; DB 10; Length 274;
Best Local Similarity 62.5%; Pred. No. 58;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 WXXXXXX 8
DB 21 WRISYMG 28

RESULT 7
Q9A956 PRELIMINARY: PRT: 613 AA.
AC Q9A956;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE TONR-DEPENDENT RECEPTOR, PUTATIVE.
GN CC1138.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
NCBI_TaxID=69394;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA Debey R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Ullerbach T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL: AE005792; AAK23122.1; -
DR TIGR: CC1138; -
KW Receptor; Complete proteome.
SQ SEQUENCE 613 AA; 64411 MW; ED65208A3D82B97D CRC64;

QY 1 WXXXXXX 8
DB 370 WRAAAYAG 377

Query Match 83.3%; Score 30; DB 2; Length 613;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

RESULT 8
Q9ET38 PRELIMINARY: PRT: 193 AA.
AC Q9ET38;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CLAUDIN-19 (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=ICR.
RA Kinichi Y., Morita K., Furuse M., Tsukita S.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF249889; AAF98323.1; -
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin. 1.
DR PRINTS: PRO1077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 193 AA; 20299 MW; 2F2D82DB5FCF0D7F CRC64;
```

```

Query Match
Best Local Similarity 80.6%; Score 29; DB 11; Length 193;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxsyxg 8
   1 111
   20 WKOSYAG 27

RESULT 9
ID 097402 PRELIMINARY; PRT; 217 AA.
AC 097402;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE XYLANASE PRECURSOR (EC 3.2.1.8).
OS Phaeon cochleariae (Mustard beetle).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cuculiformia; Phytophaga; Chrysomelidae; Chrysomelidae;
OC Chrysomelinae; Phaeon.
OX NCBI_TaxID=80249;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=GUT;
RA Girard C., Jouanin L.;
RL Submitted (MUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y17908; CAA76932.1; -.
DR HSSP; P36217; 1XVO.
DR InterPro; IPR001137; Glyco_hydro_11.
DR Pfam; PF00457; Glyco_hydro_11; 1.
DR PRINTS; PR00911; GLHYDRLASE11.
DR PROSITE; PS00776; GLYCOSYL_HYDROL_F11_1; 1.
DR PROSITE; PS00777; GLYCOSYL_HYDROL_F11_2; 1.
DR Signal; xylan degradation; Hydrolase; Glycosidase.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 217 XYLANASE.
SQ SEQUENCE 217 AA; 23942 MW; 92185C41BF39A644 CRC64;

Query Match
Best Local Similarity 80.6%; Score 29; DB 5; Length 217;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxsyxg 8
   1 111
   76 WRVYVSG 83

RESULT 10
ID 090Q06 PRELIMINARY; PRT; 479 AA.
AC 090Q06;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE PUTATIVE APHID TRANSMISSION FACTOR (FRAGMENT).
OS sugarcane yellow leaf virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
OC Unassigned Luteoviridae.
OX NCBI_TaxID=94290;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A.
RA Moonan F., Molina J.J., Mirkov T.E.;
RT "Sugarcane yellow leaf virus is a new virus with a genome that has
RT Poliovirus, Luteovirus, and Enamovirus properties.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DE EMBL; AF157029; AAD45687.1; -.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
DR InterPro; IPR002929; PLRV_ORF5.
DR InterPro; IPR002965; P_rich_extensn.

```

```

DR Pfam; PF01690; PLRV_ORF5; 1.
DR PRINTS; PR00910; LVIRUSORF6.
DR PRINTS; PR01217; PRICHEXTENS.
FT NON_TER 1
SQ SEQUENCE 479 AA; 51676 MW; 6C1DBAA2BA5EEC9 CRC64;

Query Match
Best Local Similarity 80.6%; Score 29; DB 12; Length 479;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxsyxg 8
   1 111
   129 WRVYVSG 136

RESULT 11
ID 09JH75 PRELIMINARY; PRT; 479 AA.
AC 09JH75;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE PUTATIVE APHID TRANSMISSION FACTOR (FRAGMENT).
OS sugarcane yellow leaf virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
OC Unassigned Luteoviridae.
OX NCBI_TaxID=94290;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CP65-357;
RA Smith G.R.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CP65-357;
RX MEDLINE=20318675; PubMed=10859394;
RA Smith G.R., Borg Z., Lockhart B.E.L., Brattwaite K.S., Gibbs M.J.;
RT "Sugarcane yellow leaf virus: a novel member of the Luteoviridae that
RT probably arose by inter-species recombination.";
RL J. Gen. Virol. 81:1865-1869(2000).
DR EMBL; AJ249447; CAB75437.1; -.
DR InterPro; IPR000893; Luteo_ORF6.
DR InterPro; IPR002929; PLRV_ORF5.
DR InterPro; IPR002965; P_rich_extensn.
DR Pfam; PF01690; PLRV_ORF5; 1.
DR PRINTS; PR00910; LVIRUSORF6.
DR PRINTS; PR01217; PRICHEXTENS.
FT NON_TER 1
SQ SEQUENCE 479 AA; 51616 MW; AD3B98A1B658652F CRC64;

Query Match
Best Local Similarity 80.6%; Score 29; DB 12; Length 479;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxsyxg 8
   1 111
   129 WRVYVSG 136

RESULT 12
ID 09BRR6 PRELIMINARY; PRT; 497 AA.
AC 09BRR6;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE SIMILAR TO RIKEN CDNA 2610017G09 GENE.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;

```

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-KIDNEY ADENOCARCINOMA;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC006112; AAH06112.1; -
SQ SEQUENCE 497 AA: 54088 MW: B758E977CDA88F8F CRC64;

Query Match 80.6%; Score 29; DB 4; Length 497;
Best Local Similarity 50.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxsxyxg 8
1 1 1 1
4 WRGSAYAG 11

RESULT 13
O9RKB5 PRELIMINARY; PRT; 519 AA.
AC O9RKB5
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE MONOOXYGENASE.
CN SCE87.23C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RA Seeger K.J., Harris D.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE:97000351; PubMed:8843436;
RA Redenbach M., Kleiser H.M., Denaplatte D., Elchner A., Cullum J.,
RA Kinsashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL MO1. Microbiol. 21:77-96(1996).
CC -1- COFACTOR: FAD (BY SIMILARITY).
DR EMBL: AL132674; CAB59668.1; -
DR InterPro: IPR000759; Adrxn_reductase.
DR InterPro: IPR001327; FAD_Pyr_redox.
DR InterPro: IPR000960; FMO.
DR InterPro: IPR000205; NAD_Binding.
DR InterPro: IPR00103; Pyridine_redox_2.
DR InterPro: IPR001100; Pyr_redox.
DR Pfam: PF00743; FMO-like; 1.
DR PRINTS: PRO0419; ADXRDTASE.
DR PRINTS: PRO0368; FADPNR.
DR PRINTS: PRO0370; FMOXYGENASE.
DR PRINTS: PRO0411; PNDRTASE1.
DR PRINTS: PRO0469; PNDRTASE1.
KW FAD: Flavoprotein; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 519 AA: 56673 MW: 4031FB54427A57B4 CRC64;

Query Match 80.6%; Score 29; DB 2; Length 519;
Best Local Similarity 62.5%; Pred. No. 1.8e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxsxyxg 8

Db 50 WRDNTYPG 57

RESULT 14

ID O49559 PRELIMINARY; PRT; 648 AA.

AC O49559
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE HYPOTHETICAL 73.5 KDA PROTEIN.
GN F7J7.120 OR AT4G21180.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Murphy G., Drost L., Hall C., Hudson S., Ridley P.,
RA Bancroft I., Mewes H.W., Mayer K., Scheller C.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.

RN [2]
RP SEQUENCE FROM N.A.
RA Murphy G., Ridley P., Hudson S., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL021960; CAI17537.1; -
DR EMBL: AL161554; CAB79118.1; -
DR InterPro: IPR001623; DnaJ_N.
DR Pfam: PF00226; DnaJ; 1.
DR SMART: SM00271; DnaJ; 1.
DR PROSITE: PS50076; DnaJ_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 648 AA: 73514 MW: ECFC6BAE914958D6 CRC64;

Query Match 80.6%; Score 29; DB 10; Length 648;
Best Local Similarity 50.0%; Pred. No. 2.3e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxsxyxg 8
1 1 1 1
201 WRSSKTYG 208

RESULT 15

ID O9CFH4 PRELIMINARY; PRT; 822 AA.

AC O9CFH4
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE PHOSPHOKETOLASE.
GN PTK.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN-IL403;
RA Boletín A., Wincker P., Manger S., Jallón O., Malarme K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
lactis.";
RL Genome Res. 0:0-0(2001).
DR EMBL: AE006381; AAK05600.1; -
KW Complete proteome.

SQ SEQUENCE 822 AA; 93363 MW; C686D569D3EBE22F CRC64;

Query Match
Best Local Similarity 80.6%; Score 29; DB 2; Length 822;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsyxg 8
||:|
Db 17 WRAATYLG 24

RESULT 16
Q9RPH6 PRELIMINARY; PRT; 1083 AA.

AC Q9RPH6; 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE RECB.
GN RECB.
OS Mycobacterium smegmatis.
OC Bacteria; Firmicutes; Actinobacteriia; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1772;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC2155 (NB2);
RX MEDLINE=99412429; PubMed=10481025;
RA Griffin T.J. IV, Parsons L., Leshcziner A.F., Devost J.,
Deidyshtre K.M., Grindley N.D.F.;
RT "In vitro transposition of Tn552: a tool for DNA sequencing and
mutagenesis."
RL Nucleic Acids Res. 27:3859-3865(1999).
DR EMBL: AF157643; AAD46808.1; -.
DR HSSP: p56255; 2PJR
DR InterPro: IPR001993; Mitoch.carrier.
DR InterPro: IPR00212; UvrD-helicase.
DR Pfam: PF00580; UvrD-helicase.1.
DR PROSITE: PS00215; MITOCH.CARRIER; UNKNOWN.1.
SQ SEQUENCE 1083 AA; 117442 MW; A511F5FBA5EB748 CRC64;

Query Match
Best Local Similarity 80.6%; Score 29; DB 2; Length 1083;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsyxg 8
||:|
Db 766 WKRTSYSG 773

RESULT 17
Q95978 PRELIMINARY; PRT; 157 AA.

AC Q95978; 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE VHI PROTEIN PRECURSOR (FRAGMENT).
GN VHI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PERIPHERAL BLOOD;
RA Jox A., Zander T., Kueppers R., Irsch J., Kanzler H., Kornacker M.,
Bohlen H., Diehl V., Wolf J.;
RT "Absence of immunoglobulin in Hodgkin-Reed Sternberg cells of a
patient with mixed cellularly Hodgkin's disease is associated with
somatic mutations within the untranslated regions of rearranged and

RT class switch recombined Ig genes.";
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
DOMAIN.
DR EMBL: AJ005570; CAA06599.1; -.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig; 1.
DR SMART: SM00406; IGV; 1.
DR Signal.
FT SIGNAL. 1 21
FT NON_TER 157 157 POTENTIAL.
SQ SEQUENCE 157 AA; 17304 MW; 86986EDDA84D88B5 CRC64;

Query Match
Best Local Similarity 77.8%; Score 28; DB 4; Length 157;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsyxg 8
||:|
Db 121 WRSNYNG 128

RESULT 18
Q9FHV6 PRELIMINARY; PRT; 197 AA.

AC Q9FHV6; 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE GBAAD48950.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=99397451; PubMed=10470850;
RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Kotani H.,
RA Miyajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. IX.
RT Sequence features of the regions of 1,011,550 bp covered by seventeen
RT pl and TAC clones."
RL DNA Res. 6:183-195(1999).
DR EMBL: AB017068; BAB11363.1; -.
SQ SEQUENCE 197 AA; 23374 MW; 20B0BFBF767746B7 CRC64;

Query Match
Best Local Similarity 77.8%; Score 28; DB 10; Length 197;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsyxg 8
||:|
Db 142 WRTNYHG 149

RESULT 19
Q9RBD0 PRELIMINARY; PRT; 206 AA.

AC Q9RBD0; 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 3-DEMETHYLUBIQUINONE-9-3-METHYLTRANSFERASE, PUTATIVE.
GN DR2562.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]

RP SEQUENCE FROM N.A.
 RT STRAIN=RI:
 RX MEDLINE:20036896; PubMed:10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
 RA Marwara K.S., Aravind L., Daly M.J., Minton K.W., Fleischman R.D.,
 RA Kechum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioreistant bacterium Deinococcus
 RT radiodurans R1.";
 RL Science 286:1571-1577(1999).
 DR EMBL: AE002085; AAF12100.1; .
 DR TIGR: DR2562; .
 DR InterPro: IPR001601; Meth-transf.
 DR InterPro: IPR000051; SAM bind.
 DR Transferrase; Methyltransferase; Ubiquinone; Complete proteome.
 SO SEQUENCE 206 AA; 22186 MW; 6F03E1369E12D870 CRC64;

Query Match	Score	DB 2	Length
77.88	28	206	

Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

y      1 wrxxsyxg 8
      11 1:1
b     111 wrtAsFDG 118

```

RESULT	20
ADJUDICATED	

Q9BVN0	PRELIMINARY;	PRT;	211 AA.
--------	--------------	------	---------

AC Q95VNU: 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
 DE CLAUDIN 7.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 CX NCBI_taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=CHORIOCARCINOMA;
 RA Strausberg R.;
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC001055; AAH01055.1; -
 SQ SEQUENCE 211 AA: 22420 MW: 7597DB08D3D902F0 CRC64;

Query Match	77.8%	Score 28	DB 4	Length 211
Best Similarity	50.0%	Pred. 1.2e+02		
Best Local				
Matches 4	Conservative 1	Mismatches 3	Indels 0	Gaps 0

```
QY      1 wrxxsyxg 8
db      30 wQMSsYAG 37
```

Search completed: January 14, 2002, 07:39:46
Job time: 957 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:36:00 ; Search time 103.51 Seconds
(without alignments)
5.725 Million cell updates/sec

Title: 09-185908-1E
Perfect score: 36
Sequence: 1 wrxxsyxg 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database :

_AGeneseq_1101.*
1: /SID52/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	32	88.9	8	21	AA06521
2	32	88.9	8	21	AA06574
3	32	88.9	10	21	AA06530
4	32	88.9	10	21	AA06539
5	32	88.9	10	21	AA06548
6	32	88.9	10	21	AA06557
7	32	88.9	10	21	AA06566
8	32	88.9	230	21	AA051676
9	31	86.1	8	21	AA06426
10	31	86.1	8	21	AA06512
11	31	86.1	9	21	AA06427

12	31	86.1	10	21	AA06485	Claudin-1 cyclic c
13	31	86.1	10	21	AA06491	Claudin-1 cyclic c
14	31	86.1	10	21	AA06497	Claudin-1 cyclic c
15	31	86.1	10	21	AA06503	Claudin-1 cyclic c
16	31	86.1	10	21	AA06509	Claudin-1 cyclic c
17	31	86.1	71	20	AA038421	Human secreted pro
18	31	86.1	211	20	AA041726	Human PRO944 prote
19	31	86.1	211	20	AA040413	Human Tango-73 pro
20	31	86.1	211	20	AA069653	Human senescence f
21	31	86.1	211	21	AA044282	Human PRO944 (UNO4
22	31	86.1	211	21	AA068679	A human molecule a
23	31	86.1	212	20	AA038430	Human secreted pro
24	31	86.1	212	21	AA076130	Human secreted pro
25	29	80.6	55	20	AA012227	Human 5' EST seque
26	29	80.6	56	22	AA04286	Human gene 10 enco
27	29	80.6	83	22	AA040407	Human polypeptide
28	29	80.6	113	22	AA04281	Human gene 10 enco
29	29	80.6	114	20	AA012226	Human 5' EST seque
30	29	80.6	126	22	AA025829	Human protein sequ
31	29	80.6	126	22	AA042193	Human polypeptide
32	29	80.6	140	22	AA04228	Human gene 10 enco
33	29	80.6	155	21	AA054052	Human pancreatic c
34	29	80.6	230	20	AA036134	Human secreted pro
35	29	80.6	230	20	AA036181	Human secreted pro
36	29	80.6	230	21	AA039378	Human PRO1356 (UNO
37	29	80.6	230	21	AA084609	A human membrane a
38	29	80.6	230	22	AA038857	Human polypeptide
39	29	80.6	230	22	AA012417	Human PRO1356 poly
40	29	80.6	230	22	AA04207	Human gene 10 enco
41	29	80.6	230	22	AA087565	Human PRO1356. Ho
42	29	80.6	230	22	AA088342	Human membrane or
43	29	80.6	230	22	AA066127	Protein of the inv
44	29	80.6	260	22	AA040643	Human polypeptide
45	29	80.6	448	22	AA052467	Mycobacterium tube
46	29	80.6	466	22	AA081335	Human AFP protein
47	29	80.6	466	22	AA088466	Human membrane or
48	29	80.6	467	21	AA012158	Hydrophobic domain
49	28	77.8	8	21	AA06419	Claudin-1 cell adh
50	28	77.8	8	21	AA06479	Claudin-1 cyclic c

ALIGNMENTS

RESULT 1	
AA06521	standard; peptide: 8 AA.
ID	AA06521:
AC	AA06521:
XX	
DT	28-SEP-2000 (first entry)
XX	
DE	Claudin-2 cell adhesion recognition sequence SPO ID NO: 42.
XX	
XX	Claudin-2 modulating agent; cell adhesion recognition sequence;
KW	CAR sequence: autoimmune disease; inflammatory disease; cancer;
KW	graft rejection.
XX	
OS	Mammalia.
XX	
PN	WO200026360-A1.
XX	
PD	11-MAY-2000.
XX	
PF	03-NOV-1999; 99WO-CA01029.
XX	
PR	03-NOV-1998; 98US-0185908.
XX	
PR	30-MAR-1999; 99US-0282029.
XX	
PA	(ADHE-) ADHEREX TECHNOLOGIES INC.
XX	
PT	Blaschuck OM, Symonds JM, Gour BJ:
XX	

DR WPI: 2000-365610/31.
XX Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
PS Claim 46: Page 97: 121pp; English.
XX
CC The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-2 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site.
XX
SQ Sequence 8 AA;

Query Match 88.9%; Score 32; DB 21; Length 8;
Best Local Similarity 62.5%; Pred. No. 4.3e+05;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 wrxxsyxg 8
|| || |
Db 1 wrtsyvg 8

RESULT 2
AAB06574
ID AAB06574 standard; peptide; 8 AA.
XX
AC AAB06574;
XX
DT 28-SEP-2000 (first entry)
XX
DE Claudin-2 cyclic cell adhesion recognition sequence SEQ ID NO: 163.
XX
KW Claudin-2 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection; cyclic.
XX
OS Mammalia.
XX
PN WO200026360-A1.
XX
PD 11-MAY-2000.
XX
PF 03-NOV-1999; 99WO-CA01029.
XX
PR 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
XX
PA (ADHE-) ADHEREX TECHNOLOGIES INC.
XX
PI Blaschuck OW, Symonds JM, Gour BJ;
PI WPI: 2000-365610/31.
XX
DR Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
PS Claim 49: Page 98: 121pp; English.
XX
CC The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-2 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site.

CC to the desired target site. The present sequence has a cyclic
CC conformation.
XX
SQ Sequence 8 AA;

Query Match 88.9%; Score 32; DB 21; Length 8;
Best Local Similarity 62.5%; Pred. No. 4.3e+05;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 wrxxsyxg 8
|| || |
Db 1 wrtsyvg 8

RESULT 3
AAB06530
ID AAB06530 standard; peptide; 10 AA.
XX
AC AAB06530;
XX
DT 28-SEP-2000 (first entry)
XX
DE Claudin-2 cyclic cell adhesion recognition sequence SEQ ID NO: 119.
XX
KW Claudin-2 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection; cyclic.
XX
OS Mammalia.
XX
PN WO200026360-A1.
XX
PD 11-MAY-2000.
XX
PF 03-NOV-1999; 99WO-CA01029.
XX
PR 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
XX
PA (ADHE-) ADHEREX TECHNOLOGIES INC.
XX
PI Blaschuck OW, Symonds JM, Gour BJ;
PI WPI: 2000-365610/31.
XX
DR Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
PS Claim 49: Page 98: 121pp; English.
XX
CC The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-2 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.
XX
SQ Sequence 10 AA;

Query Match 88.9%; Score 32; DB 21; Length 10;
Best Local Similarity 62.5%; Pred. No. 1.2;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 wrxxsyxg 8
|| || |
Db 2 wrtsyvg 9

RESULT	4
ID	AAB06539
AC	AAB06539 standard; peptide: 10 AA.
XX	
XX	AAB06539:
DT	28-SEP-2000 (first entry)
XX	
DE	Claudin-2 cyclic cell adhesion recognition sequence SEQ ID NO: 128.
XX	
KM	Claudin-2 modulating agent; cell adhesion recognition sequence;
KW	CAR sequence: autoimmune disease; inflammatory disease; cancer;
XX	graft rejection; cyclic.
OS	Mammalia.
PN	MO200026360-A1.
PD	11-MAY-2000.
PF	03-NOV-1999; 99WO-CA01029.
XX	
PR	03-NOV-1998; 98US-0185908.
PR	30-MAR-1999; 99US-0282029.
PA	(ADHE-) ADHEREX TECHNOLOGIES INC.
PI	Biaschuck OW, Symonds JM, Gour BJ
DR	WPI; 2000-365610/31.
PT	Antibody modulation of claudin-mediated cell adhesion for increasing
PT	vasopermeability, for delivering drugs to tumors and the nervous system
PT	and across the skin -
XX	
PS	ClaIm 49; Page 98; 121pp; English.
XX	
CC	The present invention relates to the use of peptides as claudin-mediated
CC	cell adhesion modulators. The claudin-2 group of proteins are cadherins,
CC	situations, membrane glycoproteins involved in cell adhesion. In some
CC	adhesions, cell adhesion occurs at abnormal levels, and these peptides
CC	can be used to modulate these levels, and thus treat autoimmune diseases,
CC	inflammatory diseases and cancer, and aid wound healing and implant
CC	adhesion. In addition, they can also be used to facilitate drug delivery
CC	to the desired target site. The present sequence has a cyclic
CC	conformation.
XX	
SQ	Sequence 10 AA:
Query Match	88.9%; Score 32; DB 21; Length 10;
Best Local Similarity	62.5%; Pred. NO. 1.2;
Matches 5; Conservative	0; Mismatches 3; Indels 0; Gaps 0;
OY	1 wrxxxyxg 8
Db	2 wrtasyvg 9
RESULT	5
ID	AAB06548
AC	AAB06548 standard; peptide: 10 AA.
XX	
XX	AAB06548:
DT	28-SEP-2000 (first entry)
XX	
DE	Claudin-2 cyclic cell adhesion recognition sequence SEQ ID NO: 137.
KM	Claudin-2 modulating agent; cell adhesion recognition sequence;
KW	CAR sequence: autoimmune disease; inflammatory disease; cancer;
KW	graft rejection; cyclic.

XX	OS	Mammalia.
XX	PN	WO200026360-A1.
XX	PD	11-MAY-2000.
XX	PF	03-NOV-1999; 99WO-CA01029.
XX	PR	03-NOV-1998; 98US-0185908.
XX	PR	30-MAR-1999; 99US-0282029.
XX	PA	(ADHE-) ADHEREX TECHNOLOGIES INC.
XX	PI	Blaschuck OW, Symonds JM, Gour BJ;
XX	DR	WPI: 2000-365610/31.
XX	PT	Antibody modulation of claudin-mediated cell adhesion for increasing vasopermeability, for delivering drugs to tumors and the nervous system and across the skin -
XX	PS	Claim 49; Page 98; 12Ipp; English.
XX	CC	The present invention relates to the use of peptides as claudin-mediated cell adhesion modulators. The claudin-2 group of proteins are cadherins, which are membrane glycoproteins involved in cell adhesion. In some situations, cell adhesion occurs at abnormal levels, and these peptides can be used to modulate these levels, and thus treat autoimmune diseases, inflammatory diseases and cancer, and aid wound healing and implant adhesion. In addition, they can also be used to facilitate drug delivery to the desired target site. The present sequence has a cyclic conformation.
XX	SQ	Sequence 10 AA:
OY		Query Match 88.9%; Score 32; DB 21; Length 10; Best Local Similarity 62.5%; Pred. No. 1.2; Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0.
Db		1 wrxsyxg 8 2 wrtsyvg 9
RESULT 6		
AAB06557		
ID AAB06557		standard; peptide; 10 AA.
AC AAB06557;		
XX	DT	28-SEP-2000 (first entry)
DE		Claudin-2 cyclic cell adhesion recognition sequence SEQ ID NO: 146.
KM		Claudin-2 modulating agent; cell adhesion recognition sequence;
KW		CAR sequence; autoimmune disease; inflammatory disease; cancer;
KX		graft rejection; cyclic.
OS		Mammalia.
PN		WO200026360-A1.
PD		11-MAY-2000.
PF		03-NOV-1999; 99WO-CA01029.
PR		03-NOV-1998; 98US-0185908.
PR		30-MAR-1999; 99US-0282029.
PA		(ADHE-) ADHEREX TECHNOLOGIES INC.
PI		Blaschuck OW, Symonds JM, Gour BJ;
DR		WPI: 2000-365610/31.
PT		Antibody modulation of claudin-mediated cell adhesion for increasing vasopermeability, for delivering drugs to tumors and the nervous system and across the skin -
PS		Claim 49; Page 98; 12Ipp; English.
CC		The present invention relates to the use of peptides as claudin-mediated cell adhesion modulators. The claudin-2 group of proteins are cadherins, which are membrane glycoproteins involved in cell adhesion. In some situations, cell adhesion occurs at abnormal levels, and these peptides can be used to modulate these levels, and thus treat autoimmune diseases, inflammatory diseases and cancer, and aid wound healing and implant adhesion. In addition, they can also be used to facilitate drug delivery to the desired target site. The present sequence has a cyclic conformation.
SQ		Sequence 10 AA:

PI Blaschuck OW, Symonds JM, Gour BJ;
XX WPI; 2000-365610/31.
XX
XX Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
XX
PS Claim 49; Page 98; 121pp; English.
XX
CC The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-2 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.
CC
SQ Sequence 10 AA;

Query Match 88.9%; Score 32; DB 21; Length 10;
Best Local Similarity 62.5%; Pred. No. 1.2;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxsyxg 8
|| || |
Db 2 wrtssyvg 9

RESULT 7
AAB06566
ID AAB06566 standard; peptide; 10 AA.
XX
AC AAB06566;
XX
DT 28-SEP-2000 (first entry)
XX
DE claudin-2 cyclic cell adhesion recognition sequence SEQ ID NO: 155.
XX
XX claudin-2 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection; cyclic.
XX
OS Mammalia.
XX
XX WO200026360-A1.
XX
PD 11-MAY-2000.
XX
PF 03-NOV-1999; 99WO-CA01029.
XX
XX 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
XX
XX (ADHE-) ADHEREX TECHNOLOGIES INC.
PA
XX Blaschuck OW, Symonds JM, Gour BJ;
PI WPI; 2000-365610/31.
DR
XX
PT Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
XX
PS Claim 49; Page 98; 121pp; English.
XX
CC The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-2 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides

CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.
CC
SQ Sequence 10 AA;

Query Match 88.9%; Score 32; DB 21; Length 10;
Best Local Similarity 62.5%; Pred. No. 1.2;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxsyxg 8
|| || |
Db 2 wrtssyvg 9

RESULT 8
AAV51676
ID AAV51676 standard; Protein; 230 AA.
XX
AC AAV51676;
XX
DT 02-JUN-2000 (first entry)
XX
DE Murine clodin 2 protein.
XX
XX Clodin 2; murine; tight junction-constituting membrane protein;
KW medicine.
XX
OS Mus sp.
XX
PN JP2000032984-A.
XX
PD 02-FEB-2000.
XX
PF 26-JUN-1998; 98JP-0179847.
XX
PR 15-MAY-1998; 98JP-0133215.
XX
PA (EISA) EISAI CO LTD.
XX
XX WPI: 2000-285512/25.
DR N-PSDB; AA89137.
XX
XX Tight junction-constituting membrane protein clodin family - useful in
PT the medical field
PT
PS Claim 2; Page 10; 22pp; Japanese.
XX
CC This invention describes novel murine nucleic acid sequences encoding the
CC clodin family of tight junction (TJ)-constituting membrane protein. The
CC membrane protein can be used in medical field. This sequence represents
CC the clodin 2 protein described in the method of the invention.
XX
SQ Sequence 230 AA;

Query Match 88.9%; Score 32; DB 21; Length 230;
Best Local Similarity 62.5%; Pred. No. 25;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxsyxg 8
|| || |
Db 30 wrtssyvg 37

RESULT 9
AAB06426
ID AAB06426 standard; peptide; 8 AA.
XX
AC AAB06426;

XX 28-SEP-2000 (first entry)
XX
XX Claudin-1 cell adhesion recognition sequence SEQ ID NO: 485.
DE
XX Claudin-1 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection.
XX
XX Mammalia.
OS
XX WO200026360-A1.
PN
XX 11-MAY-2000.
PD
XX
XX 03-NOV-1999; 99WO-CA01029.
PF
XX 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
XX
XX (ADHE-) ADHEREX TECHNOLOGIES INC.
PA
XX Blaschuck OW, Symonds JM, Gour BJ;
PI
XX WPI; 2000-365610/31.
DR
XX Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
XX Claim 39; Page 96; 121pp; English.
XX
XX The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-1 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site.
XX
XX Sequence 8 AA;
SQ

Query Match 86.1%; Score 31; DB 21; Length 8;
Best Local Similarity 62.5%; Pred. No. 4.3e+05;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 wrxxsyxg 8
II III
DB 1 wrlsyxg 8

RESULT 10
AAB06512
ID AAB06512 standard; peptide; 8 AA.
XX
XX AAB06512;
AC
XX 28-SEP-2000 (first entry)
XX
XX Claudin-1 cyclic cell adhesion recognition sequence SEQ ID NO: 519.
DE
XX Claudin-1 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection; cyclic.
XX
XX Mammalia.
OS
XX WO200026360-A1.
PN
XX 11-MAY-2000.
PD
XX

PF 03-NOV-1999; 99WO-CA01029.
XX
XX 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
XX
XX (ADHE-) ADHEREX TECHNOLOGIES INC.
PA
XX Blaschuck OW, Symonds JM, Gour BJ;
PI
XX WPI; 2000-365610/31.
DR
XX
XX Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
XX Claim 43; Page 97; 121pp; English.
XX
XX The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-1 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.
XX
XX Sequence 8 AA;
SQ

Query Match 86.1%; Score 31; DB 21; Length 8;
Best Local Similarity 62.5%; Pred. No. 4.3e+05;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 wrxxsyxg 8
II III
DB 1 wrlsyxg 8

RESULT 11
AAB06427
ID AAB06427 standard; peptide; 9 AA.
XX
XX AAB06427;
AC
XX 28-SEP-2000 (first entry)
XX
XX Claudin-1 cell adhesion recognition sequence SEQ ID NO: 486.
DE
XX Claudin-1 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection.
XX
XX Mammalia.
OS
XX WO200026360-A1.
PN
XX 11-MAY-2000.
PD
XX
XX 03-NOV-1999; 99WO-CA01029.
PF
XX 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
XX
XX (ADHE-) ADHEREX TECHNOLOGIES INC.
PA
XX Blaschuck OW, Symonds JM, Gour BJ;
PI
XX WPI; 2000-365610/31.
DR
XX
XX Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX

XX Claim 39; Page 96; 121pp; English.

XX The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-1 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site.

XX Sequence 9 AA;

Query Match 86.1%; Score 31; DB 21; Length 9;
Best Local Similarity 62.5%; Pred. No. 4.3e+05;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsyxg 8
|| || |
Db 1 wrlsyag 8

RESULT 12

AAB06485
ID AAB06485 standard; peptide; 10 AA.

AC AAB06485;

DT 28-SEP-2000 (first entry)

DE Claudin-1 cyclic cell adhesion recognition sequence SEQ ID NO: 492.

KW Claudin-1 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection; cyclic.

XX Mammalia.

XX WO200026360-A1.

XX 11-MAY-2000.

XX 03-NOV-1999; 99WO-CA01029.

XX 03-NOV-1998; 98US-0185908.

XX 30-MAR-1999; 99US-0282029.

XX (ADHE-) ADHEREX TECHNOLOGIES INC.

XX Blaschuck OM, Symonds JM, Gour BJ;

XX WPI; 2000-365610/31.

XX Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -

XX Claim 43; Page 96; 121pp; English.

XX The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-1 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.

XX Sequence 10 AA;

Query Match 86.1%; Score 31; DB 21; Length 10;
Best Local Similarity 62.5%; Pred. No. 1.9;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsyxg 8
|| || |
Db 2 wrlsyag 9

RESULT 13

AAB06491
ID AAB06491 standard; peptide; 10 AA.

AC AAB06491;

DT 28-SEP-2000 (first entry)

DE Claudin-1 cyclic cell adhesion recognition sequence SEQ ID NO: 498.

KW Claudin-1 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection; cyclic.

XX Mammalia.

XX WO200026360-A1.

XX 11-MAY-2000.

XX 03-NOV-1999; 99WO-CA01029.

XX 03-NOV-1998; 98US-0185908.

XX 30-MAR-1999; 99US-0282029.

XX (ADHE-) ADHEREX TECHNOLOGIES INC.

XX Blaschuck OM, Symonds JM, Gour BJ;

XX WPI; 2000-365610/31.

XX Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -

XX Claim 43; Page 96; 121pp; English.

XX The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-1 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.

XX Sequence 10 AA;

Query Match 86.1%; Score 31; DB 21; Length 10;
Best Local Similarity 62.5%; Pred. No. 1.9;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsyxg 8
|| || |
Db 2 wrlsyag 9

RESULT 14

AAB06497
ID AAB06497 standard; peptide; 10 AA.

AC AAB06497;
XX 28-SEP-2000 (first entry)
XX
DE Claudin-1 cyclic cell adhesion recognition sequence SEQ ID NO: 504.
XX
XX Claudin-1 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
XX graft rejection; cyclic.
XX Mammalia.
OS
XX WO200026360-A1.
PN
XX
XX 11-MAY-2000.
XX
XX 03-NOV-1999; 99WO-CA01029.
PF
XX 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
XX
XX (ADHE-) ADHEREX TECHNOLOGIES INC.
PA
XX Blaschuck OW, Symonds JM, Gour BJ;
PI WPI: 2000-365610/31.
DR
XX
XX Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
XX
PS Claim 43; Page 97; 121pp; English.
XX
XX The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-1 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.
XX
XX Sequence 10 AA;
SO

Query Match 86.1%; Score 31; DB 21; Length 10;
Best Local Similarity 62.5%; Pred. No. 1.9;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxsyxg 8
| | | | |
DB 2 wrlysyag 9

RESULT 15
AAB06503
ID AAB06503 standard; peptide; 10 AA.
XX
XX AAB06503;
AC
XX 28-SEP-2000 (first entry)
XX
XX
XX Claudin-1 cyclic cell adhesion recognition sequence SEQ ID NO: 510.
DE
XX
XX Claudin-1 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection; cyclic.
XX Mammalia.
OS
XX WO200026360-A1.
PN
XX

PD 11-MAY-2000.
XX
XX
PF 03-NOV-1999; 99WO-CA01029.
XX
XX 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
XX
XX (ADHE-) ADHEREX TECHNOLOGIES INC.
PA
XX Blaschuck OW, Symonds JM, Gour BJ;
PI WPI: 2000-365610/31.
DR
XX
XX Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
XX
PS Claim 43; Page 97; 121pp; English.
XX
XX The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-1 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.
XX
XX Sequence 10 AA;
SO

Query Match 86.1%; Score 31; DB 21; Length 10;
Best Local Similarity 62.5%; Pred. No. 1.9;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxsyxg 8
| | | | |
DB 2 wrlysyag 9

RESULT 16
AAB06509
ID AAB06509 standard; peptide; 10 AA.
XX
XX AAB06509;
AC
XX 28-SEP-2000 (first entry)
XX
XX
XX Claudin-1 cyclic cell adhesion recognition sequence SEQ ID NO: 516.
DE
XX
XX Claudin-1 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection; cyclic.
XX Mammalia.
OS
XX WO200026360-A1.
PN
XX
XX 11-MAY-2000.
XX
XX 03-NOV-1999; 99WO-CA01029.
PF
XX 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
XX
XX (ADHE-) ADHEREX TECHNOLOGIES INC.
PA
XX Blaschuck OW, Symonds JM, Gour BJ;
PI WPI: 2000-365610/31.
DR
XX
XX Antibody modulation of claudin-mediated cell adhesion for increasing
PT

PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
XX
XX Claim 43; Page 97; 121pp; English.
XX
XX The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-1 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.
XX
SQ Sequence 10 AA;

Query Match 86.1%; Score 31; DB 21; Length 10;
Best Local Similarity 62.5%; Pred. No. 1.9;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxsyxg 8
11 11 1
Db 2 wriysyag 9

RESULT 17
AY38421
ID AAY38421 standard; Protein; 71 AA.
XX
XX AAY38421;
DT 30-SEP-1999 (first entry)
XX
XX Human secreted protein encoded by gene No. 36.
DE
XX
XX Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumor; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischemic shock; Alzheimer's disease; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
PN
XX WO935158-A1.
XX
PD 15-JUL-1999.
XX
PF 06-JAN-1999; 99WO-US00108.
XX
XX 07-JAN-1998; 98US-0070704.
PR 07-JAN-1998; 98US-0070657.
PR 07-JAN-1998; 98US-0070658.
PR 07-JAN-1998; 98US-0070692.
XX
XX (HUMA-) HUMAN GENOME SCT INC.
PA
PI Brewer LA, Duan RD, Ebner R, Lafleur DW, Ni J;
PI Olsen HS, Rosen CA, Ruben SM, Soppet DR;
XX
XX WPI; 1999-444190/37.
DR N-PSDB; AA206254.
XX
XX New isolated human genes and the secreted polypeptides they encode
PS Claim 11; Page 195; 227pp; English.
XX
XX This sequence represents a secreted human protein encoded by the gene
CC clone detailed in the descriptor line. The gene can be used to generate

CC fusion proteins by linking to the gene to a human immunoglobulin Fc
CC portion (e.g. AA06210) for increasing the stability of the fused
CC protein as compared to the human protein only.
CC The invention relates to 36 novel genes and their fragments (nucleic
CC acid sequences: AA06219-206263; amino acid sequences AAY38386-Y38498)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 36
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AA06219 for described uses).
XX
SQ Sequence 71 AA;

Query Match 86.1%; Score 31; DB 20; Length 71;
Best Local Similarity 62.5%; Pred. No. 13;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxsyxg 8
11 11 1
Db 30 wriysyag 37

RESULT 18
AAY1726
ID AAY1726 standard; Protein; 211 AA.
XX
XX AAY1726;
AC
XX
DT 07-DEC-1999 (first entry)
XX
XX Human PRO944 protein sequence.
DE
XX
XX Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW secreted protein; transmembrane protein.
XX
OS Homo sapiens.
PN
XX WO946281-A2.
XX
PD 16-SEP-1999.
XX
PF 08-MAR-1999; 99WO-US05028.
XX
XX 10-MAR-1998; 98US-0077450.
PR 11-MAR-1998; 98US-0077632.
PR 11-MAR-1998; 98US-0077641.
PR 11-MAR-1998; 98US-0077649.
PR 12-MAR-1998; 98US-0077791.
PR 13-MAR-1998; 98US-0078004.
PR 17-MAR-1998; 98US-0040220.
PR 20-MAR-1998; 98US-0078886.
PR 20-MAR-1998; 98US-0078910.
PR 20-MAR-1998; 98US-0078936.
PR 20-MAR-1998; 98US-0078939.
PR 25-MAR-1998; 98US-0079294.
PR 26-MAR-1998; 98US-0079656.
PR 27-MAR-1998; 98US-0079663.
PR 27-MAR-1998; 98US-0079664.
PR 27-MAR-1998; 98US-0079689.
PR 27-MAR-1998; 98US-0079728.
PR 27-MAR-1998; 98US-0079786.
PR 30-MAR-1998; 98US-0079920.
PR 30-MAR-1998; 98US-0079923.
PR 31-MAR-1998; 98US-0080105.
PR 31-MAR-1998; 98US-0080107.
PR 31-MAR-1998; 98US-0080165.
PR 31-MAR-1998; 98US-0080194.
PR 01-APR-1998; 98US-0080327.
PR 01-APR-1998; 98US-0080328.

PT	adhesion disorders -	98US-008333.	01-APR-1998.
PR		98US-0080334.	01-APR-1998.
PR		98US-0081049.	08-APR-1998.
PR		98US-0081070.	08-APR-1998.
PR		98US-0081071.	08-APR-1998.
PR		98US-0081195.	09-APR-1998.
PR		98US-0081203.	09-APR-1998.
PR		98US-0081229.	09-APR-1998.
PR		98US-0081817.	15-APR-1998.
PR		98US-0081838.	15-APR-1998.
PR		98US-0081952.	15-APR-1998.
PR		98US-0081955.	15-APR-1998.
PR		98US-0082568.	21-APR-1998.
PR		98US-0082569.	21-APR-1998.
PR		98US-0082700.	22-APR-1998.
PR		98US-0082804.	22-APR-1998.
PR		98US-0082804.	23-APR-1998.
PR		98US-0082767.	23-APR-1998.
PR		98US-0082796.	27-APR-1998.
PR		98US-0083336.	27-APR-1998.
PR		98US-0083392.	28-APR-1998.
PR		98US-0083392.	29-APR-1998.
PR		98US-0083445.	29-APR-1998.
PR		98US-0083554.	29-APR-1998.
PR		98US-0083558.	29-APR-1998.
PR		98US-0083559.	29-APR-1998.
PR		98US-0083742.	30-APR-1998.
PR		98US-0084366.	05-MAY-1998.
PR		98US-0084414.	06-MAY-1998.
PR		98US-0084598.	07-MAY-1998.
PR		98US-0084600.	07-MAY-1998.
PR		98US-0084627.	07-MAY-1998.
PR		98US-0084637.	07-MAY-1998.
PR		98US-0084639.	07-MAY-1998.
PR		98US-0084640.	07-MAY-1998.
PR		98US-0084643.	13-MAY-1998.
PR		98US-0085323.	13-MAY-1998.
PR		98US-0085338.	13-MAY-1998.
PR		98US-0085339.	15-MAY-1998.
PR		98US-0085573.	15-MAY-1998.
PR		98US-0085579.	15-MAY-1998.
PR		98US-0085580.	15-MAY-1998.
PR		98US-0085582.	15-MAY-1998.
PR		98US-0085689.	15-MAY-1998.
PR		98US-0085697.	15-MAY-1998.
PR		98US-0085700.	15-MAY-1998.
PR		98US-0085704.	18-MAY-1998.
PR		98US-0086023.	22-MAY-1998.
PR		98US-0086392.	22-MAY-1998.
PR		98US-0086414.	22-MAY-1998.
PR		98US-0086430.	22-MAY-1998.
PR		98US-0086486.	22-MAY-1998.
PR		98US-0087098.	28-MAY-1998.
PR		98US-0087106.	28-MAY-1998.
PR		98US-0087208.	28-MAY-1998.
PR		98US-0094651.	30-JUL-1998.
PR		98US-0100038.	11-SEP-1998.
XX			
PA	(GETH) GENENTECH INC.		
XX			
PI	Wood WI, Coddard A, Gurney A, Yuan J, Baker KP, Chen J;		
XX			
DR	WPI: 1999-551358/46.		
DR	N-PSDB; AA234118.		
XX			
PT	Now secreted and transmembrane polypeptides and their polynucleotides useful for treating blood coagulation disorders, cancers and cellular		

CC	The present invention describes secreted and transmembrane polypeptides
CC	and their polynucleotides. The nucleotide sequences are useful as
CC	sources of probes, primers, for chromosome mapping, and for generation
CC	of antisense sequences. They can also be used to create transgenic
CC	animals. The proteins can be used to treat a variety of diseases and
CC	disorders, depending on their function. Diseases that may be treated
CC	include blood coagulation disorders, cancers and cellular adhesion
CC	disorder. They may also be used to raise antibodies. AA23891 to
CC	AA234338, and AAY41685 to AAY41774 represent polynucleotide and
CC	polypeptide sequence given in the exemplification of the present
CC	invention.
XX	
XX	
SQ	Sequence 211 AA;
XX	
XX	
Query Match	86.1%; Score 31; DB 20; Length 211;
Best Local Similarity	62.5%; Pred. No. 37;
Matches 5; Conservative	0; Mismatches 3; Indels 0; Gaps 0;
OY	1 WRRXSXSG 8
Dd	30 WRYSYSG 37
XX	
RESULT 19	
AAV04143	
ID	AAV04143 standard; Protein: 211 AA.
XX	
AC	AAV04143;
XX	
DT	15-JUN-1999 (first entry)
DE	Human Tango-73 protein.
XX	
KM	Human; Tango-71; Tango-73; Tango-74; Tango-76; Tango-83; diagnosis;
KW	detection.
XX	
OS	Homo sapiens.
XX	
PN	WO9907850-A1.
PD	18-FEB-1999.
XX	
PF	06-AUG-1998; 98MO-US16502.
XX	
PR	05-SEP-1997; 97US-0058108.
PR	06-AUG-1997; 97US-0054966.
XX	
PA	(MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
XX	
I1	Goodearl ADJ, Holtzman DA;
XX	
DR	WI: 1999-167426/14.
DR	N-PsDB; AAX19956.
XX	
PT	New TANGO polypeptides and nucleic acids encoding them - useful as
PT	diagnostic agents and for treating disorders caused by aberrant
PT	expression of TANGO
XX	
PS	Claim 8; Fig 2; 8app; English.
XX	
..	The present sequence represents human Tango-73. Tango polypeptides are
CC	useful for identifying compounds which bind the polypeptide via direct
CC	binding, competition binding assays or Tango-71, -73, -74, 76 or -83-
CC	mediated signal transduction. Tango polypeptides are also useful for
CC	identifying modulating compounds by determining effect on Tango activity.
CC	Tango polypeptides and nucleic acids are useful for diagnosing diseases
CC	related to aberrant expression of Tango, and Tango polypeptides are
CC	useful for raising antibodies which can be used in diagnostic assays for
CC	detection of Tango, and also for generating anti-idiotypic antibodies for

CC prevention and protection.
XX
SQ Sequence 211 AA;

Query Match 86.1%; Score 31; DB 20; Length 211;
Best Local Similarity 62.5%; Pred. No. 37;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 wrxxsyxg 8
|| || |
DB 30 wrlysyag 37

RESULT 20

AAW99653
ID AAW99653 standard; Protein; 211 AA.
XX
AC AAW99653;
XX
XX
DT 21-MAY-1999 (first entry)
DE Human senescence factor p23 protein.
XX
XX Human; senescence factor; p23; cancer; persistent inflammation;
KW proliferative disorder; degenerative disorder.
XX
OS Homo sapiens.
XX
PN W09907893-A1.
XX
PD 18-FEB-1999.
XX
PF 05-AUG-1998; 98WO-US16343.
XX
PR 08-AUG-1997; 97US-0908873.
XX
PA (UNIW) UNIV WASHINGTON.
XX
PI Hosier S, Kublies M, Swissheim K;
XX
DR WPI; 1999-167454/14.
XX
DR N-PSDB; AAX19461.
XX
PT Newly isolated nucleic acid molecule (designated p23) encoding a p23
XX polypeptide - useful for inducing a senescence phenotype in a cell
XX
PS Claim 3; Page 35; 44pp; English.

The present sequence is human senescence factor p23. An expression
vector for p23 is useful for inducing a senescent phenotype in a cell
(preferably eukaryotic). This may help in regulating diseases, including
cancer, persistent inflammation, and various proliferative and
degenerative disorders. These transgenic cells are useful in gene
therapy for treating cancer, particularly where antisense
oligonucleotides are useful for blocking normal or mutant p23 expression
in cancer cells or other proliferating cells. Transgenic cells are also
useful for producing the p23 polypeptide in large quantities. The
antibodies are useful for raising antiserum against p23; and for
identifying senescent cells in culture and tissue biopsies. The p23
polynucleotides are useful for modulating or altering p23 activity in a
cell, and for identifying and isolating the whole gene encoding p23,
and variants of p23. Assays based on p23 elements, which detect p23
levels and activity are useful as diagnostic markers for staging tumours,
determining prognosis, and/or predicting therapeutic success. These
elements also provide an assay for detecting chromosomal rearrangements
in chromosome 3 in a human cell. The isolation of the p23 polynucleotide
permits the manipulation of malignant growth in cancer.

XX Sequence 211 AA;

Query Match 86.1%; Score 31; DB 20; Length 211;

Best Local Similarity 62.5%; Pred. No. 37;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 wrxxsyxg 8
|| || |
DB 30 wrlysyag 37

Search completed: January 14, 2002, 07:36:01
Job time: 812 sec

,

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on : January 14, 2002, 07:23:43 ; Search time 48.05 Seconds
(without alignments)
3.747 Million cell updates/sec

Title: 09-185908-1E
Perfect score: 36
Sequence: 1 wrxxsyxg 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/Backfilest1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	28	77.8	1289	1	US-07-876-280-4
2	28	77.8	1289	1	US-07-675-772-4
3	28	77.8	1289	1	US-08-063-170-4
4	28	77.8	1289	1	US-08-158-232-4
5	28	77.8	1289	1	US-08-304-626-4
6	28	77.8	1289	1	US-08-316-301A-4
7	28	77.8	1289	2	US-08-611-928-4
8	28	77.8	1289	3	US-09-173-891-4
9	28	77.8	1289	4	US-09-076-137-4
10	28	77.8	1289	5	PCT-US92-03624-4
11	27	75.0	748	1	US-08-408-318-2
12	27	75.0	748	1	US-08-369-796-10
13	27	75.0	748	1	US-08-839-164-2
14	27	75.0	748	2	US-08-852-091-10
15	27	75.0	748	2	US-08-820-754-10
16	27	75.0	748	3	US-08-956-652-10
17	27	75.0	748	3	US-08-956-869-10
18	27	75.0	748	3	US-08-948-547-10
19	27	75.0	748	4	US-09-087-465-8
20	27	75.0	748	5	PCT-US95-17025-10
21	27	75.0	749	1	US-08-276-099A-15
22	27	75.0	749	1	US-08-781-890-15
23	27	75.0	749	3	US-09-012-710-9
24	27	75.0	749	4	US-09-364-970-4
25	27	75.0	851	1	US-08-369-796-2
26	27	75.0	851	2	US-08-852-091-2
27	27	75.0	851	2	US-08-820-754-2

28	27	75.0	851	3	US-08-956-652-2	Sequence 2, Appl1
29	27	75.0	851	3	US-08-956-869-2	Sequence 2, Appl1
30	27	75.0	851	3	US-09-012-710-2	Sequence 2, Appl1
31	27	75.0	851	3	US-08-948-547-2	Sequence 2, Appl1
32	27	75.0	851	4	US-09-087-465-4	Sequence 4, Appl1
33	27	75.0	851	4	US-09-364-970-2	Sequence 2, Appl1
34	27	75.0	851	5	PCT-US95-17025-2	Sequence 2, Appl1
35	27	75.0	852	1	US-08-276-099A-13	Sequence 13, Appl1
36	27	75.0	852	1	US-08-781-890-13	Sequence 13, Appl1
37	27	75.0	1220	1	US-08-158-232-43	Sequence 43, Appl1
38	27	75.0	1220	2	US-08-611-928-43	Sequence 43, Appl1
39	27	75.0	1220	3	US-09-173-891-43	Sequence 43, Appl1
40	27	75.0	1289	6	5281530-3	Patent No. 5281530
41	27	75.0	1289	6	5426049-4	Patent No. 5426049
42	27	75.0	1385	1	US-07-876-280-2	Sequence 2, Appl1
43	27	75.0	1385	1	US-07-675-772-2	Sequence 2, Appl1
44	27	75.0	1385	1	US-08-063-170-2	Sequence 2, Appl1
45	27	75.0	1385	1	US-08-158-232-2	Sequence 2, Appl1
46	27	75.0	1385	1	US-08-304-626-2	Sequence 2, Appl1
47	27	75.0	1385	1	US-08-316-301A-2	Sequence 2, Appl1
48	27	75.0	1385	2	US-08-611-928-2	Sequence 2, Appl1
49	27	75.0	1385	3	US-09-173-891-2	Sequence 2, Appl1
50	27	75.0	1385	4	US-09-076-137-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-07-876-280-4
Sequence 4, Application US/07876280
Patent No. 5262158
GENERAL INFORMATION:
APPLICANT: Payne, Jewel M.
APPLICANT: Cannon, Raymond J.C.
TITLE OF INVENTION: No. 5262158el Bacillus thuringiensis Isolates for
TITLE OF INVENTION: Controlling Acarides
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/876,280
FILING DATE: 19920430
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: M/S 104
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1289 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: BACILLUS THURINGIENSIS

INDIVIDUAL ISOLATE: PS17
IMMEDIATE SOURCE:
CLONE: E. coli NM522(pMYC 1628) NRRL B-18652
US-07-876-280-4

Query Match 77.8%; Score 28; DB 1; Length 1289;
Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxsxyxg 8
DB 382 WRAQYGG 389

RESULT 2
US-07-675-772-4
Sequence 4, Application US/07675772
Patent No. 5262399
GENERAL INFORMATION:

APPLICANT: Hickie, Leslie A.
APPLICANT: Sick, August J.
APPLICANT: Schwab, George E.
APPLICANT: Narva, Kenneth E.
APPLICANT: Payne, Jewel M.
TITLE OF INVENTION: No. 5262399el Compositions and Methods for the control of
TITLE OF INVENTION: Flukes
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROMAN SALIMANCHIK
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/675,772
FILING DATE: 19910327
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SALIMANCHIK, ROMAN
REGISTRATION NUMBER: 21,023
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1289 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: BACILLUS THURINGIENSIS
INDIVIDUAL ISOLATE: PS17
IMMEDIATE SOURCE:
LIBRARY: LAMBDAGEM (TM) - 11 LIBRARY OF KENNETH NARVA
CLONE: 17B
US-07-675-772-4

Query Match 77.8%; Score 28; DB 1; Length 1289;
Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxsxyxg 8

DB 382 WRAQYGG 389

RESULT 3
US-08-063-170-4
Sequence 4, Application US/08063170
Patent No. 5350576
GENERAL INFORMATION:

APPLICANT: Kim, Leo
APPLICANT: Schwab, George E.
TITLE OF INVENTION: Compositions and Methods for Inducing an Immune
TITLE OF INVENTION: Response for Protection Against Endoparasites and Exoparas
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: DAVID R. SALIMANCHIK
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/063,170
FILING DATE: 19930517
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,141
FILING DATE: 30-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/759,248
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: SALIMANCHIK, DAVID R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: M/S 103.C2
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1289 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus thuringiensis
STRAIN: PS17
INDIVIDUAL ISOLATE: PS17B
US-08-063-170-4

Query Match 77.8%; Score 28; DB 1; Length 1289;
Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxsxyxg 8
DB 382 WRAQYGG 389

RESULT 4
US-08-158-232-4
Sequence 4, Application US/08158232
Patent No. 5596071
GENERAL INFORMATION:

APPLICANT: Payne, Jewel
APPLICANT: Kennedy, M. Keith
APPLICANT: Randall, John Brooks
APPLICANT: Meier, Henry
APPLICANT: Vick, Heidi Jane
APPLICANT: Foncecrada, Luis
APPLICANT: Schnepf, H. Ernest
APPLICANT: Schwab, George E.
APPLICANT: Fu, Jenny
TITLE OF INVENTION: No. 5596071el Bacillus thuringiensis Toxins Active
TITLE OF INVENTION: Against Hymenopteran Pests
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/158,232
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/887,980
FILING DATE: 22-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/797,645
FILING DATE: 25-NOV-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/703,977
FILING DATE: 22-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: M/SCJ104.C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1289 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: BACILLUS THURINGIENSIS
STRAIN: PS17
IMMEDIATE SOURCE:
CLONE: E. coli NM522(pMYC1628) NRRL B-18652
US-08-158-232-4
Query Match 77.8%; Score 28; DB 1; Length 1289;
Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 1 wrxxsyxg 8
|| |
Db 382 WRAOYCG 389
RESULT 5
US-08-304-626-4

Sequence 4, Application US/08304626
Patent No. 5616495
GENERAL INFORMATION:
APPLICANT: Payne, Jewel M.
APPLICANT: Kennedy, M. Keith
APPLICANT: Randall, John Brooks
APPLICANT: Meier, Henry
APPLICANT: Vick, Heidi Jane
APPLICANT: Foncecrada, Luis
APPLICANT: Schnepf, Harry E.
APPLICANT: Schwab, George E.
TITLE OF INVENTION: No. 5616495el Bacillus thuringiensis Isolates
TITLE OF INVENTION: Active Against Hymenopteran Pests and Genes Encoding
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/304,626
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/887,980
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: M/SCJ 104
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1289 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: BACILLUS THURINGIENSIS
STRAIN: PS17
IMMEDIATE SOURCE:
CLONE: E. coli NM522(pMYC1628) NRRL B-18652
US-08-304-626-4
Query Match 77.8%; Score 28; DB 1; Length 1289;
Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 1 wrxxsyxg 8
|| |
Db 382 WRAOYCG 389
RESULT 6
US-08-316-301A-4
Sequence 4, Application US/08316301A
Patent No. 5753492
GENERAL INFORMATION:

APPLICANT: Schnepf, Harry E.
APPLICANT: Schwab, George E.
APPLICANT: Payne, Jewel M.
APPLICANT: Narva, Kenneth E.
APPLICANT: Poncetrada, Luis
TITLE OF INVENTION: No. 5753492e1 Nematode-Active Toxins and Genes
TITLE OF INVENTION: 42
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/316,301A
FILING DATE: 30-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/871,510
FILING DATE: 23-APR-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/693,018
FILING DATE: 03-MAY-1991
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/565,544
FILING DATE: 10-AUG-1990
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/084,653
FILING DATE: 12-AUG-1987
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/830,050
FILING DATE: 31-JAN-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lloyd, Jeff
REGISTRATION NUMBER: 35,589
REFERENCE/DOCKET NUMBER: MA20CCDD1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1289 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: BACILLUS THURINGIENSIS
INDIVIDUAL ISOLATE: PS17
IMMEDIATE SOURCE:
CLONE: E. coli NM522(pMYC 1628) NRRL B-18652
US-08-316-301A-4

Query Match 77.8%; Score 28; DB 1; Length 1289;
Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 1 wrxsxyxg 8

Db 382 WRAQYGC 389
RESULT 7
US-08-611-928-4
Sequence 4, Application US/08611928
Patent No. 5824792
GENERAL INFORMATION:
APPLICANT: Payne, Jewel
APPLICANT: Kennedy, M. Keith
APPLICANT: Randall, John Brooks
APPLICANT: Meier, Henry
APPLICANT: Uick, Heidi Jane
APPLICANT: Poncetrada, Luis
APPLICANT: Schnepf, H. Ernest
APPLICANT: Schwab, George E.
APPLICANT: Fu, Jenny
TITLE OF INVENTION: No. 5824792e1 Bacillus thuringiensis Toxins Active
TITLE OF INVENTION: Against Hymenopteran Pests
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/611,928
FILING DATE: 06-MAR-1996
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/158,232
FILING DATE: 24-NOV-1993
APPLICATION NUMBER: US 07/887,980
FILING DATE: 22-MAY-1992
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/797,645
FILING DATE: 25-NOV-1991
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/703,977
FILING DATE: 22-MAY-1991
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: M/SCJ104.C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-372-5800
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1289 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: BACILLUS THURINGIENSIS
INDIVIDUAL ISOLATE: PS17b
IMMEDIATE SOURCE:
CLONE: E. coli NM522(pMYC1628) NRRL B-18652

US-08-611-928-4

Query Match 77.8%; Score 28; DB 2; Length 1289;
Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxxsyxg 8
DB 382 WRAAOYCG 389

RESULT 8
US-09-173-891-4
Sequence 4, Application US/09173891
Patent No. 6077937

GENERAL INFORMATION:

APPLICANT: Payne, Jewel
APPLICANT: Kennedy, M. Keith
APPLICANT: Randall, John Brooks
APPLICANT: Meier, Henry
APPLICANT: Ulick, Heidi Jane
APPLICANT: Foncerra, Luis
APPLICANT: Schepf, H. Ernest
APPLICANT: Schwab, George E.
APPLICANT: Fu, Jenny
TITLE OF INVENTION: No. 6077937e1 Bacillus thuringiensis Toxins Active
TITLE OF INVENTION: Against Hymenopteran Pests
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSER: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/173,891
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/158,232

FILING DATE: US 07/887,980
FILING DATE: 22-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/797,645
FILING DATE: 25-NOV-1991
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/703,977
FILING DATE: 22-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: W/SCJ104.C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1289 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:

ORGANISM: BACILLUS THURINGIENSIS
STRAIN: PS17
INDIVIDUAL ISOLATE: PS17b
IMMEDIATE SOURCE:
CLONE: E. coli NM522(pMVC1628) NRRL B-18652
US-09-173-891-4

Query Match 77.8%; Score 28; DB 3; Length 1289;
Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxxsyxg 8
DB 382 WRAAOYCG 389

RESULT 9
US-09-076-137-4
Sequence 4, Application US/09076137B
Patent No. 6166195

GENERAL INFORMATION:
APPLICANT: Schepf, Harry E.
APPLICANT: Payne, Jewel M.
APPLICANT: Schwab, George E.
APPLICANT: Naiva, Kenneth E.
APPLICANT: Foncerra, Luis
TITLE OF INVENTION: No. 6166195e1 Nematode-Active Toxins and Genes Which Code
TITLE OF INVENTION: Therefor
FILE REFERENCE: MA-20CCCD2
CURRENT APPLICATION NUMBER: US/09/076,137B
CURRENT FILING DATE: 1998-05-12
EARLIER APPLICATION NUMBER: 08/316,301
EARLIER FILING DATE: 1994-09-30
NUMBER OF SEQ ID NOS: 42
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 4
LENGTH: 1289
TYPE: PRT
ORGANISM: Bacillus thuringiensis
US-09-076-137-4

Query Match 77.8%; Score 28; DB 4; Length 1289;
Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxxsyxg 8
DB 382 WRAAOYCG 389

RESULT 10
PCT-US92-03624-4
Sequence 4, Application PC/TUS9203624

GENERAL INFORMATION:
APPLICANT: Schepf, Harry E.
APPLICANT: Schwab, George E.
APPLICANT: Payne, Jewel M.
APPLICANT: Naiva, Kenneth E.
APPLICANT: Foncerra, Luis
TITLE OF INVENTION: Novel Nematode-Active Toxins and Genes
TITLE OF INVENTION: Which Code Therefor
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSER: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/03624
FILING DATE: 19920501
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Salimanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: MA20C2C1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1289 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: BACILLUS THURINGIENSIS
INDIVIDUAL ISOLATE: PS17
IMMEDIATE SOURCE:
CLONE: E. coli NM522(pMYC 1628) NRRL B-18652
PCT-US92-03624-4

Query Match 77.8%; Score 28; DB 5; Length 1289;
Best Local Similarity 50.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxxsyxg 8
DB 382 WRAQYGG 389

RESULT 11
US-08-408-318-2
Sequence 2, Application US/08408318
Patent No. 5639858
GENERAL INFORMATION:
APPLICANT: Hoev, Timothy
TITLE OF INVENTION: Human Signal Transducers and Binding
TITLE OF INVENTION: Assays
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herber
STREET: 850 Hansen Way, #200
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/408,318
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A.
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-60845
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-494-8700
TELEFAX: 415-494-8771
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 748 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-408-318-2

Query Match 75.0%; Score 27; DB 1; Length 748;
Best Local Similarity 50.0%; Pred. No. 3.4e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsyxg 8
DB 500 WQFSSYVG 507

RESULT 12
US-08-369-796-10
Sequence 10, Application US/08369796
Patent No. 5716622
GENERAL INFORMATION:
APPLICANT: James E. Darnell, Jr.
APPLICANT: Zilong Wen
APPLICANT: Curt M. Horvath
APPLICANT: Zhong Zhong
TITLE OF INVENTION: FUNCTIONALLY ACTIVE REGIONS OF SIGNAL
TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION (STAT) PROTEINS
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/369,796
FILING DATE: 06-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-116
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 748 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-369-796-10

Query Match 75.0%; Score 27; DB 1; Length 748;
Best Local Similarity 50.0%; Pred. No. 3.4e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsyxg 8
DB 500 WQFSSYVG 507

RESULT 13
US-08-839-164-2

Sequence 2, Application US/08839164
Patent No. 5756700
GENERAL INFORMATION:
APPLICANT: Hoey, Timothy
TITLE OF INVENTION: Human Signal Transducers and Binding
TITLE OF INVENTION: Assays
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herber
STREET: 850 Hansen Way, #200
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,164
FILING DATE: 23-APR-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/408,318
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A.
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-60845
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-494-8700
TELEFAX: 415-494-8771
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 748 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-839-164-2

Query Match 75.0%; Score 27; DB 1; Length 748;
Best Local Similarity 50.0%; Pred. No. 3.4e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxsxyxg 8
I: | | |
Db 500 WQSSYVG 507

RESULT 14
US-08-852-091-10
Sequence 10, Application US/08852091
Patent No. 5883228
GENERAL INFORMATION:
APPLICANT: James E. Darnell, Jr.
APPLICANT: Zilong Wen
APPLICANT: Curt M. Horvath
APPLICANT: Zhong Zhong
TITLE OF INVENTION: FUNCTIONALLY ACTIVE REGIONS OF SIGNAL
TITLE OF INVENTION: TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION (STAT) PROTEINS
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/852,091
FILING DATE: 06-MAY-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/369,796
FILING DATE: 06-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-116
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 748 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-852-091-10

Query Match 75.0%; Score 27; DB 2; Length 748;
Best Local Similarity 50.0%; Pred. No. 3.4e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxsxyxg 8
I: | | |
Db 500 WQSSYVG 507

RESULT 15
US-08-820-754-10
Sequence 10, Application US/08820754
Patent No. 5976835
GENERAL INFORMATION:
APPLICANT: Darnell Jr., James E.
APPLICANT: Schindler, Christian W.
APPLICANT: Fu, Xian-Yuan
APPLICANT: Wen, Zilong
APPLICANT: Zhong, Zhong
TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN
TITLE OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/820,754
FILING DATE: 19-MAR-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/212,185
FILING DATE: 11-MAR-1994
APPLICATION NUMBER: US 07/980,498
FILING DATE: 23-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/854,296
FILING DATE: 19-MAR-1992
PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO US93/02569
FILING DATE: 19-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/126,588
FILING DATE: 24-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-073 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 748 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-820-754-10

Query Match 75.0%; Score 27; DB 2; Length 748;
Best Local Similarity 50.0%; Pred. No. 3.4e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsyxg 8
I: | | |
Db 500 WQFSSYVG 507

RESULT 16
US-08-956-652-10
Sequence 10, Application US/08956552
Patent No. 6013475
GENERAL INFORMATION:
APPLICANT: Darnell Jr., James E.
APPLICANT: Schindler, Christian W.
APPLICANT: Fu, Xian-Yuan
APPLICANT: Wen, Zilong
APPLICANT: Zhong, Zhong
TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN
TITLE OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,652
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/212,185
FILING DATE: 11-MAR-1994
APPLICATION NUMBER: US 07/980,498
FILING DATE: 23-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/854,296
FILING DATE: 19-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO US93/02569
FILING DATE: 19-MAR-1993
PRIOR APPLICATION DATA: US 08/126,588

FILING DATE: 24-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-073 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 748 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-956-652-10

Query Match 75.0%; Score 27; DB 3; Length 748;
Best Local Similarity 50.0%; Pred. No. 3.4e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsyxg 8
I: | | |
Db 500 WQFSSYVG 507

RESULT 17
US-08-956-869-10
Sequence 10, Application US/08956869
Patent No. 6030808
GENERAL INFORMATION:
APPLICANT: Darnell Jr., James E.
APPLICANT: Schindler, Christian W.
APPLICANT: Fu, Xian-Yuan
APPLICANT: Wen, Zilong
APPLICANT: Zhong, Zhong
TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN
TITLE OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,869
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/212,185
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/854,296
FILING DATE: 19-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO US93/02569
FILING DATE: 19-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/126,588
FILING DATE: 24-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-073 CIP
TELECOMMUNICATION INFORMATION:

TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 748 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-956-869-10

Query Match 75.0%; Score 27; DB 3; Length 748;
Best Local Similarity 50.0%; Pred. No. 3.4e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxsyxg 8
1: 111
Db 500 WQSSSYVG 507

RESULT 18
US-08-948-547-10
Sequence 10, Application US/08948547
Patent No. 6124118
GENERAL INFORMATION:
APPLICANT: Darnell Jr., James E.
APPLICANT: Schlindler, Christian W.
APPLICANT: Fu, Xian-Yuan
APPLICANT: Wen, Zilong
APPLICANT: Zhong, Zhong
TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN
NUMBER OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/948,547
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/212,185
FILING DATE: 11-MAR-1994
APPLICATION NUMBER: US 07/980,498
FILING DATE: 23-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/854,296
FILING DATE: 19-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO US93/02569
FILING DATE: 19-MAR-1993
APPLICATION DATA:
APPLICATION NUMBER: US 08/126,588
FILING DATE: 24-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-073 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:
LENGTH: 748 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-948-547-10

Query Match 75.0%; Score 27; DB 3; Length 748;
Best Local Similarity 50.0%; Pred. No. 3.4e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxsyxg 8
1: 111
Db 500 WQSSSYVG 507

RESULT 19
US-09-087-465-8
Sequence 8, Application US/09087465A
Patent No. 6160092
GENERAL INFORMATION:
APPLICANT: Vinkemeier, Uwe
APPLICANT: Chen, Xiaomin
APPLICANT: Darnell Jr., James E.
APPLICANT: Kuriyan, John
TITLE OF INVENTION: A CRYSTAL OF THE CORE PORTION OF A STAT AND METHODS OF
FILE REFERENCE: 600-1-229
CURRENT APPLICATION NUMBER: US/09/087,465A
NUMBER OF SEQ ID NOS: 37
CURRENT FILING DATE: 1998-05-29
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 8
LENGTH: 748
TYPE: PRT
ORGANISM: Homo sapiens
US-09-087-465-8

Query Match 75.0%; Score 27; DB 4; Length 748;
Best Local Similarity 50.0%; Pred. No. 3.4e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxsyxg 8
1: 111
Db 500 WQSSSYVG 507

RESULT 20
PCT-US95-17025-10
Sequence 10, Application PC/TUS9517025
GENERAL INFORMATION:
APPLICANT: James E. Darnell, Jr.
APPLICANT: Zilong Wen
APPLICANT: Curt M. Horvath
APPLICANT: Zhong Zhong
TITLE OF INVENTION: FUNCTIONALLY ACTIVE REGIONS OF SIGNAL
TITLE OF INVENTION: TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION (STAT) PROTEINS
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/17025
FILING DATE: 28-DEC-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/369,796
FILING DATE: 06-JAN-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-116
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 748 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-17025-10

Query Match 75.0%; Score 27; DB 5; Length 748;
Best local Similarity 50.0%; Pred. No. 3.4e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 wrxxsyxg 8
1: 111
DB 500 WQFSSYVG 507

Search completed: January 14, 2002, 07:23:44
Job time: 75 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:37:29 : Search time 63.57 Seconds
(Without alignments)
9.586 Million cell updates/sec

Title: 09-185908-1f
Perfect score: 36
Sequence: 1 wrxxayxg 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	31	86.1	513	1	A35742	aqualysin (EC 3.4.
2	30	83.3	260	2	F82120	zinc ABC transport
3	30	83.3	261	2	A64066	probable membrane
4	30	83.3	262	2	F82959	permease of ABC 21
5	30	83.3	303	2	T00479	probable phosphati
6	29	80.6	199	1	TJBPXL	tail assembly prot
7	29	80.6	224	2	B85584	probable tail comp
8	29	80.6	448	2	H70947	hypothetical prote
9	29	80.6	1180	1	NCEKX5	exodeoxyribonuclea
10	29	80.6	1180	2	G85933	hypothetical prote
11	29	80.6	1208	2	B82091	exodeoxyribonuclea
12	28	77.8	146	2	E72167	A32L protein - var
13	28	77.8	146	2	J01834	3L protein - var
14	28	77.8	146	2	S46858	A31L protein - var
15	28	77.8	278	2	H71119	hypothetical prote
16	28	77.8	303	2	J01386	hypothetical 33k p
17	28	77.8	551	2	T11657	hypothetical prote
18	28	77.8	648	2	T04949	hypothetical prote
19	28	77.8	832	2	F86812	phosphoketolase (l
20	27	75.0	154	2	E81795	probable tRNA/rRNA
21	27	75.0	154	2	E81218	RNA methyltransfer
22	27	75.0	159	2	S41178	gene 36 protein -
23	27	75.0	168	2	S21041	cytochrome-c oxida
24	27	75.0	175	2	T11179	NADH dehydrogenase
25	27	75.0	192	2	B70352	hypothetical prote
26	27	75.0	246	2	T37169	hypothetical prote
27	27	75.0	261	2	G64948	probable membrane
28	27	75.0	261	2	E85798	hypothetical prote
29	27	75.0	280	2	A39484	androgen-withdrawa

30	27	75.0	330	2	T34972	probable membrane
31	27	75.0	362	2	S69698	ERD1 protein - yea
32	27	75.0	370	2	B83191	alcohol dehydrogen
33	27	75.0	531	2	T49058	hypothetical prote
34	27	75.0	538	2	F83354	probable sulfatase
35	27	75.0	988	2	H71336	conserved hypotet
36	27	75.0	1289	2	T18212	paraspotal crystal
37	26	72.2	90	2	B81035	hypothetical prote
38	26	72.2	100	2	F84231	hypothetical prote
39	26	72.2	133	1	CCOFCT	cytochrome c' - Rh
40	26	72.2	148	2	T31141	transcription regu
41	26	72.2	154	2	H83160	hypothetical prote
42	26	72.2	160	2	E64158	hypothetical prote
43	26	72.2	179	2	A33164	hypothetical prote
44	26	72.2	188	2	D75125	gmp synthase, ater
45	26	72.2	189	2	D71006	probable GMP synth
46	26	72.2	198	2	G85506	hypothetical prote
47	26	72.2	211	2	B71058	hypothetical prote
48	26	72.2	216	2	T49795	related to 5-oxoac
49	26	72.2	224	2	G75087	transcription regu
50	26	72.2	238	2	T51072	hypothetical prote

ALIGNMENTS

RESULT 1
A35742
aqualysin (EC 3.4.21.-) I precursor - Thermus aquaticus
C:Species: Thermus aquaticus
C:Date: 10-Sep-1999 #sequence-revision 10-Sep-1999 #extl-change 16-Jun-2000
C:Accession: A35742; S00620; S00324
R:Terada, I.; Kwon, S.-T.; Miyata, Y.; Matsuzawa, H.; Ohta, T.
J. Biol. Chem. 265, 6576-6581, 1990
A:Title: Unique precursor structure of an extracellular protease, aqualysin I, with N
A:Reference number: A35742; MUID:90216674
A:Accession: A35742
A:Molecule type: DNA
A:Residues: 1-513 <TER>
A:Cross-references: GB:J90108; GB:D90108; GB:J05414; NID:9217171; PIDN:BA14135.1; PI
R:Kwon, S.-T.; Terada, I.; Matsuzawa, H.; Ohta, T.
Eur. J. Biochem. 173, 491-497, 1988
A:Title: Nucleotide sequence of the gene for aqualysin I (a thermophilic alkaline sc
A:Reference number: S00620; MUID:88225062
A:Accession: S00620
A:Molecule type: DNA
A:Residues: 75-442 <KWO>
A:Cross-references: EMBL:X07734; NID:g48069; PIDN:CA30559.1; PID:6602091
R:Matsuzawa, H.; Tokugawa, K.; Hamackl, M.; Mizoguchi, M.; Taguchi, H.; Terada, I.; K
Eur. J. Biochem. 171, 441-447, 1988
A:Title: Purification and characterization of aqualysin I (a thermophilic alkaline sc
A:Reference number: S00324; MUID:88151937
A:Accession: S00324
A:Molecule type: protein
A:Residues: 128-170 <MATS>
C:Superfamily: subtilisin; subtilisin homology
C:Keywords: extracellular protein; hydrolase; serine proteinase
F:1-14/Domain: signal sequence #status predicted <SIG>
F:15-127/Domain: propeptide #status predicted <PRO>
F:128-408/Product: aqualysin I #status experimental <MAT>
F:157-364/Domain: subtilisin homology <SBT>
F:255-257,281-283/Region: SI specificity crevice #status predicted
F:409-513/Domain: carboxyl-terminal propeptide #status predicted <CPR>
F:166,197,349/Active site: Asp, His, Ser #status predicted

Query Match 86.1%; Score 31; DB 1; Length 513;
Best Local Similarity 62.5%; Pred. No. 17;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 wrxxayxg 8

Db 495 WRKYAYSG 502

RESULT 2

F82120
zinc ABC transporter, permease protein VC2083 [Imported] - Vibrio cholerae (strain NI696)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
A:Accession: F82120
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
Chaisson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoli, I.; Sellers, F.
I.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: AB2035; MUID:20406833
A:Accession: F82120
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-260 <HEI>
A:Cross-references: GB:AE004282; GB:AE003852; NID:g9656626; PIDN:AAF95229.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain NI6961; biotype El Tor
C:Genetics:
A:Gene: VC2083
A:Map position: 1
C:Superfamily: conserved hypothetical protein HI0360

Query Match 83.3%; Score 30; DB 2; Length 260;
Best Local Similarity 62.5%; Pred. No. 15;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
| | | | |
Db 28 WRKMYFG 35

RESULT 3

A64066
Probable membrane protein HI0407 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
A:Accession: A64066
R:Feilschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
J.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
D.M.; Brandon, R.C.; Fine, L.D.; Fitchman, J.L.; Fuhmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Guelim, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630
A:Accession: A64066
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-261 <TIG>
A:Cross-references: GB:U32724; GB:LA2023; NID:g1573378; PIDN:AAC22066.1; PID:g1573380; T
C:Superfamily: conserved hypothetical protein HI0360
C:Keywords: transmembrane protein; transport protein
F:4-20/Domain: transmembrane #status predicted <TM1>
F:38-54/Domain: transmembrane #status predicted <TM2>
F:58-74/Domain: transmembrane #status predicted <TM4>
F:88-104/Domain: transmembrane #status predicted <TM5>
F:127-143/Domain: transmembrane #status predicted <TM6>
F:168-184/Domain: transmembrane #status predicted <TM7>
F:186-202/Domain: transmembrane #status predicted <TM8>
F:218-224/Domain: transmembrane #status predicted <TM9>
F:240-256/Domain: transmembrane #status predicted <TM9>

Query Match 83.3%; Score 30; DB 2; Length 261;
Best Local Similarity 62.5%; Pred. No. 15;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8

Db 28 WRKMYFG 35

RESULT 4

F82959
permease of ABC zinc transporter znub PA5501 [Imported] - Pseudomonas aeruginosa (str
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
A:Accession: F82959
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Iarbig, K.; L.
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: AB2950; MUID:20437337
A:Accession: F82959
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-262 <STO>
A:Cross-references: GB:AE004962; GB:AE004091; NID:g9951826; PIDN:AAC08886.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: znub; PA5501
C:Superfamily: conserved hypothetical protein HI0360

Query Match 83.3%; Score 30; DB 2; Length 262;
Best Local Similarity 62.5%; Pred. No. 15;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
| | | | |
Db 28 WRKMYFG 35

RESULT 5

T00479
Probable phosphatidylinositol-glycan synthase [Imported] - Arabidopsis thaliana
N:Alternate names: hypothetical protein F1913.21
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 23-Mar-2001
A:Accession: T00479; B84763
R:Rounsley, S.D.; Lin, X.; Kethum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K
submitted to the EMBL data library, April 1998
A:Description: Arabidopsis thaliana chromosome II BAC F1913 genomic sequence.
A:Reference number: Z14160
A:Accession: T00479
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-303 <ROU>
A:Cross-references: EMBL:AC004238; NID:g3033373; PID:g3033393
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffatt, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: AB4420; MUID:20083487
A:Accession: B84763
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-303 <STO>
A:Cross-references: GB:AE002093; NID:g3033393; PIDN:AAC12837.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g34980; F1913.21
A:Map position: 2

Query Match 83.3%; Score 30; DB 2; Length 303;
Best Local Similarity 62.5%; Pred. No. 18;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
11 111
DB 15 WRKAYGCG 22

RESULT 6

tail assembly protein K - phage lambda
C:Species: phage lambda
C>Date: 13-Jun-1983 #sequence_revision 13-Jun-1983 #text_change 23-Jul-1999
C:Accession: H43009; G43013; A04355
R:Danilels, D.
submitted to the Nucleic Acid Sequence Database, September 1982
A:Reference number: A94614
A:Accession: H43009
A:Molecule type: DNA
A:Residues: 1-199 <DNA>
R:Sanger, F.; Coulson, A.R.; Hong, G.F.; Hill, D.F.; Petersen, G.B.
J. Mol. Biol. 162, 729-773, 1982
A:Title: Nucleotide sequence of bacteriophage lambda DNA.
A:Reference number: A92891; MUID:83189071
A:Accession: G43013
A:Molecule type: DNA
A:Residues: 1-199 <SANG>
A:Cross-references: GB:J02459; GB:M17233; GB:M24325; GB:V00636; GB:X00906; NID:g215104;
A:Note: there are two possible initiation sites for gene K translation, the codon for 1-
C:Comment: Gene K protein is involved in the assembly of the initiator complex for tail
C:Genetics:
A:Gene: K
A:Map position: 29.43-30.66
C:Superfamily: phage lambda tail assembly protein K

Query Match 80.6%; Score 29; DB 1; Length 199;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
11 111
DB 180 WRASAFYTG 187

RESULT 7

probable tail component of prophage CP-933K 20978 [Imported] - Escherichia coli (strain
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: B85584
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Mitter, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimmlanta, E.; Potamousis, K.; Apodaca,
Natura 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B85584
A:Status: preliminary;
A:Molecule type: DNA
A:Residues: 1-224 <STO>
A:Cross-references: GB:AE005174; NID:g12513746; PIDN:AAG55134.1; GSPDB:GM00145; UMGP:209
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: 20978
C:Superfamily: phage lambda tail assembly protein K

Query Match 80.6%; Score 29; DB 2; Length 224;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
11 111
DB 205 WRASAFYTG 212

RESULT 8
H70947
hypothetical protein RV3170 - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: H70947
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Scares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:96295987
A:Accession: H70947
A:Status: preliminary;
A:Molecule type: DNA
A:Residues: 1-448 <COL>
A:Cross-references: GB:AL021646; GB:AL123456; NID:g3242278; PIDN:CAA16635.1; PID:e124
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: RV3170

Query Match 80.6%; Score 29; DB 2; Length 448;
Best Local Similarity 50.0%; Pred. No. 44;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
11 111
DB 306 WRASGYSG 313

RESULT 9

NCEC35
exodeoxyribonuclease V (EC 3.1.11.5) 135K chain - Escherichia coli
N:Alternate names: exonuclease 135K polypeptide; recB DNase 135K polypeptide
C:Species: Escherichia coli
C>Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 19-Jan-2001
C:Accession: A25532; E65064
R:Finch, P.W.; Storey, A.; Chapman, K.E.; Brown, K.; Hickson, I.D.; Emmerson, P.T.
Nucleic Acids Res. 14, 8573-8582, 1986
A:Title: Complete nucleotide sequence of the Escherichia coli recB gene.
A:Reference number: A25532; MUID:87066729
A:Accession: A25532
A:Molecule type: DNA
A:Residues: 1-1180 <FIN>
A:Cross-references: GB:X04581; NID:g42680; PIDN:CAA28250.1; PID:g42682
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A:Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: E65064
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1180 <BLAT>
A:Cross-references: GB:AE000365; GB:U00096; NID:g2367163; PIDN:AAC75859.1; PID:g17891
A:Experimental source: strain K-12, substrain MG1655
C:Comment: This enzyme is required for efficient DNA repair: it catalyzes the unwinding
of these activities require concomitant hydrolysis of ATP.
C:Genetics:
A:Gene: recB
A:Map position: 61 min
C:Superfamily: exodeoxyribonuclease V 135K chain
C:Keywords: Amp; DNA repair; hydrolase; nucleotide binding; P-loop
P:23-30/Region: nucleotide-binding motif A (P-loop)

Query Match 80.6%; Score 29; DB 1; Length 1180;
Best Local Similarity 50.0%; Pred. No. 1180;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8

```
Db      901 WRVTSYSG 908
      11 : 11
      855933
RESULT  10
hypothetical protein recB [Imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: G85933
R:Perina, N.T.; Plummett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
  11ler, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoustis, K.; Apodaca,
  Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: G85933
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1180 <STO>
A:Cross-references: GB:AE005174; NID:912517302; PIDN:AGS57931.1; GSPDB:GN00145; UMGF:Z41
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: recB
C:Superfamily: exodeoxyribonuclease V 135k chain

Query Match      80.6%; Score 29; DB 2; Length 1180;
Best Local Similarity 50.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY      1 wrxxayxg 8
      11 : 11
      901 WRVTSYSG 908

RESULT  11
exodeoxyribonuclease V, 135 kDa chain VC2320 [Imported] - Vibrio cholerae (strain N16961
882091
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: B82091
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
  chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, F
  1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
  Nature 406, 477-483, 2000
A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: B82091
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1208 <HEI>
A:Cross-references: GB:AE004303; GB:AE003852; NID:9656890; PIDN:AAF95464.1; GSPDB:GN001
C:Genetics:
A:Gene: VC2320
A:Map position: 1
C:Superfamily: exodeoxyribonuclease V 135k chain

Query Match      80.6%; Score 29; DB 2; Length 1208;
Best Local Similarity 50.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY      1 wrxxayxg 8
      11 : 11
      906 WRVTSYSG 913

RESULT  12
E72167
A32L protein - variola minor virus (strain Garcia-1966)
C:Species: variola minor virus
C:Date: 24-Nov-1999 #sequence_revision 24-Nov-1999 #text_change 20-Jun-2000
```

```
C:Accession: E72167
R:Shchelkunov, S.N.; Totmenin, A.V.; Gutorov, V.V.; Safonov, P.F.; Massung, R.F.; Lo
  submitted to Genbank, March 1998
A:Description: Analysis of the complete coding sequence of DNA of alastrim variola mi
  A:Reference number: A72150
A:Accession: E72167
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-146 <SHC>
A:Cross-references: GB:Y16780; NID:95830555; PIDN:CAB54735.1; PID:95830696
A:Experimental source: strain Garcia-1966
C:Genetics:
A:Gene: A32L
C:Superfamily: sheep pox virus HM3 protein

Query Match      77.8%; Score 28; DB 2; Length 146;
Best Local Similarity 62.5%; Pred. No. 25;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY      1 wrxxayxg 8
      11 : 11
      73 WRCVATPG 80

RESULT  13
J01834
3L protein - variola major virus
C:Species: variola major virus
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 21-Jul-2000
C:Accession: J01834; I28573
R:Aguado, B.; Selmes, I.P.; Smith, G.L.
  J. Gen. Virol. 73, 2887-2902, 1992
A:Title: Nucleotide sequence of 21.8 kbp of variola major virus strain Harvey and com
  A:Reference number: J01832; MUID:93057361
A:Accession: J01834
A:Molecule type: DNA
A:Residues: 1-146 <AGU>
A:Experimental source: strain Harvey
R:Massung, R.F.; Esposito, J.J.; Liu, L.I.; Qi, J.; Utterback, T.R.; Knight, J.C.; Au
  Nature 366, 748-751, 1993
A:Title: Potential virulence determinants in terminal regions of variola smallpox vir
  A:Reference number: Z20488; MUID:94088747
A:Accession: T28573
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-146 <MAS>
A:Cross-references: EMBL:U22579; NID:9623595; PIDN:AAA60883.1; PID:9623596
A:Experimental source: strain Bangladesh 1975
C:Superfamily: sheep pox virus HM3 protein

Query Match      77.8%; Score 28; DB 2; Length 146;
Best Local Similarity 62.5%; Pred. No. 25;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY      1 wrxxayxg 8
      11 : 11
      73 WRCVATPG 80

RESULT  14
S46858
A31L protein - variola virus
N:Alternate names: A28L protein
C:Species: variola virus
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 23-Mar-2001
C:Accession: S46858; E36851
R:Kolykhalov, A.A.; Bilnov, V.M.; Frolov, I.V.; Totmenin, A.V.; Shchelkunov, S.N.; Sa
  submitted to the EMBL Data Library, April 1992
A:Description: Nucleotide sequence analysis of the region of Variola virus xhoI-E gen
  A:Reference number: S46841
A:Accession: S46858
```

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-146 <K0L>
A:Cross-references: EMBL:X67115; NID:g516412; PIDN:CAA47502.1; PID:g516418
A:Experimental source: strain India-1967, isolate Ind3
R:Blinov, V.M.
submitted to GenBank, November 1992
A:Reference number: A36859
A:Accession: E36851
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-146 <BLI>
A:Cross-references: GB:X69198; NID:g456758; PIDN:CAA49076.1; PID:g457026
A:Experimental source: strain India1967, ssp. major, isolate Ind3
C:Superfamily: Sheep pox virus HM3 protein

Query Match 77.8%; Score 28; DB 2; Length 146;
Best Local Similarity 62.5%; Pred. No. 25;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
|| || |
Db 73 WRGVAYPG 80

RESULT 15
H71119
hypothetical protein PH0727 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 05-Nov-1999
A:Accession: H71119
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
R.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A>Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; MID:98344137
A:Accession: H71119
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-278 <KAM>
A:Cross-references: GB:AP000003; NID:g3236130; PIDN:BA429818.1; PID:dl030761; PID:g32571
A:Experimental source: strain OT3
A:Note: This accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH0727

Query Match 77.8%; Score 28; DB 2; Length 278;
Best Local Similarity 50.0%; Pred. No. 46;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
|| || |
Db 99 WRVSYWG 106

RESULT 16
J01386
hypothetical 33K protein - pea enation mosaic virus
C:Species: pea enation mosaic virus, PEVY
C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 08-Oct-1999
A:Accession: J01386
R:Demler, S.A.; de Zoeten, G.A.
J. Gen. Virol. 72, 1819-1824, 1991
A>Title: The nucleotide sequence and nucleovirus-like nature of RNA 1 of an aphid non-tr
A:Reference number: J01382; MID:91341468
A:Accession: J01386
A:Molecule type: genomic RNA
A:Residues: 1-303 <DEM>
A:Cross-references: GB:I04573; NID:g294105; PIDN:AAA72298.1; PID:g294107
A:Experimental source: strain MSG
A:Note: this reading frame extends between two stop codons and does not begin with a std

A:Note: 65-Asp was also found
C:Genetics:
A:Map position: segment RNA1

Query Match 77.8%; Score 28; DB 2; Length 303;
Best Local Similarity 50.0%; Pred. No. 50;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
|| || |
Db 135 WQARAYTG 142

RESULT 17
T16557
hypothetical protein K04E7.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
A:Accession: T16557
R:Nhan, M.
submitted to the EMBL Data Library, October 1995
A:Description: The sequence of C. elegans cosmid K04E7.
A:Reference number: Z18535
A:Accession: T16557
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-551 <NHA>
A:Cross-references: EMBL:U39666; NID:g1049408; PID:g1049411; PIDN:AAA80412.1; CESP:K0
C:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:K04E7.3
A:Introns: 43/3; 81/3; 112/3; 153/1; 191/2; 213/3; 241/2; 261/3; 324/1; 364/1; 393/3;

Query Match 77.8%; Score 28; DB 2; Length 551;
Best Local Similarity 50.0%; Pred. No. 91;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
|| || |
Db 376 WRNYSYSG 383

RESULT 18
T04949
hypothetical protein F7J7.120 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
A:Accession: T04949
R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.
submitted to the Protein Sequence Database, July 1998
A:Reference number: Z15391
A:Accession: T04949
A:Molecule type: DNA
A:Residues: 1-648 <BEV>
A:Cross-references: EMBL:AL021960
A:Experimental source: cultivar Columbia; BAC clone F7J7
C:Genetics:
A:Map position: 4
A:Introns: 44/3; 76/3; 128/3; 174/3; 194/1; 269/3; 329/3; 358/3
A:Note: F7J7.120

Query Match 77.8%; Score 28; DB 2; Length 648;
Best Local Similarity 50.0%; Pred. No. 11e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
|| || |
Db 201 WRSSKYTG 208

```
RESULT 19
F86812
phosphoketolase [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 23-Mar-2001
C:Accession: F86812
R:Boilotin, A.; Wincker, P.; Mauger, S.; Jallou, O.; Malarme, K.; Weissenbach, J.; Ehrlich
Genome Res. in press, 2001
A:Title: The complete genome sequence of the lactic acid bacterium.
A:Reference number: A86625
A:Accession: F86812
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-822 <STO>
A:Cross-references: GB:AE005176; NID:g12724499; PIDN:AK05600.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: plk

Query Match 77.8%; Score 28; DB 2; Length 822;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 wrxxayxg 8
   || |
   17 WRATATYLG 24

Db 17 WRATATYLG 24

RESULT 20
E81795
Probable tRNA/rRNA methyltransferase NMA2218 [imported] - Neisseria meningitidis (strain
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: E81795
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holtroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; MUID:20222556
A:Accession: E81795
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-154 <PAR>
A:Cross-references: GB:AL162758; GB:AL157959; NID:g7380672; PIDN:CAB85429.1; PID:g738083
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: NMA2218
C:Superfamily: Chlamyidophila pneumoniae RNA methylase

Query Match 75.0%; Score 27; DB 2; Length 154;
Best Local Similarity 50.0%; Pred. No. 44;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 wrxxayxg 8
   || |
   145 WRQHGYAG 152

Db 145 WRQHGYAG 152

Search completed: January 14, 2002, 07:37:30
Job time: 900 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OK protein - protein search, using sw model

Run on: January 14, 2002, 07:40:39 : Search time 37.71 Seconds
(without alignments)
7.778 Million cell updates/sec

Title: 09-185908-1f
Perfect score: 36
Sequence: 1 wrxxxyxx 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	31	86.1	513 1 AQL1_THERAO	P08594 thermus aqu
2	31	86.1	934 1 HMDH_PICJA	O74164 pichia fadl
3	30	83.3	261 1 ZNUB_HAETN	P44691 haemophilus
4	29	80.6	199 1 VTAK_LAMB	P03729 bacterioph
5	29	80.6	230 1 CLD2_MOUSE	O88552 mus muscul
6	29	80.6	454 1 AOFH_MYCTU	O53320 mycobacteri
7	29	80.6	1180 1 EX5B_ECOLI	P08394 escherichia
8	28	77.8	146 1 VA28_VARV	P33847 variola vir
9	28	77.8	211 1 CLD1_HUMAN	O95832 homo sapien
10	28	77.8	215 1 CLD3_BRARE	O97892 brachydanio
11	28	77.8	551 1 YVD3_CAREL	P55114 caenorhabd
12	28	77.8	822 1 XFP_LACIA	O96f4 lactococcus
13	28	77.8	1018 1 HIRA_CHICK	P79987 gallus gall
14	27	75.0	159 1 VAT_CAMVP	P18818 cauliflowe
15	27	75.0	168 1 OOX2_SUIAC	P39479 sulfobus
16	27	75.0	209 1 CLD4_CERAE	O19005 cercopithe
17	27	75.0	209 1 CLD4_HUMAN	O14493 homo sapien
18	27	75.0	209 1 CLD2_BRARE	O97890 brachydanio
19	27	75.0	210 1 CLD4_MOUSE	O35054 mus muscul
20	27	75.0	219 1 CLD3_MOUSE	O97009 mus muscul
21	27	75.0	219 1 CLD3_RAT	O63400 rattus norv
22	27	75.0	220 1 CLD3_HUMAN	O15551 homo sapien
23	27	75.0	224 1 CLDH_HUMAN	P56750 homo sapien
24	27	75.0	261 1 ZNUB_ECOLI	P39832 escherichia
25	27	75.0	362 1 ERD1_YEAST	P15151 saccharomy
26	27	75.0	773 1 ADL1_MOUSE	O97154 mus muscul
27	27	75.0	1289 1 CSAB_BACUD	O45753 bacillus th
28	26	72.2	133 1 CYCP_RHOTE	P00153 rhodocyclu
29	26	72.2	160 1 YTBK_HARIN	P44868 haemophilus
30	26	72.2	230 1 CLD2_HUMAN	P57739 homo sapien
31	26	72.2	239 1 CLDE_HUMAN	O95500 homo sapien
32	26	72.2	239 1 CLDE_MOUSE	O95083 mus muscul
33	26	72.2	269 1 COX3_EMENT	P00421 emeticella

ALIGNMENTS

RESULT 1	ID	STANDARD:	PRT:	513 AA.
AC	AQL1_THERAO			
AC	P08594:			
DT	01-AUG-1988 (Rel. 08, Created)			
DT	01-AUG-1990 (Rel. 15, Last sequence update)			
DT	01-NOV-1995 (Rel. 32, Last annotation update)			
DE	AQUALYSIN I PRECURSOR (EC 3.4.21.-).			
GN	PSIT.			
OS	Thermus aquaticus.			
OC	Bacteria; Thermus/Delnococcus group; Thermus group; Thermus.			
OX	NCBI_Taxid=271;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 15-23.			
RC	STRAIN-YT1;			
RX	MEDLINE=90216674; PubMed=2182621;			
RA	Terada I., Kwon S.-T., Miyata Y., Matsuzawa H., Ohta T.;			
RT	"Unique precursor structure of an extracellular protease, aqualysin I, with NH2- and COOH-terminal pro-sequences and its processing in Escherichia coli.";			
RL	J. Biol. Chem. 265:6576-6581(1990).			
RN	[2]			
RP	SEQUENCE OF 75-442 FROM N.A., AND PARTIAL SEQUENCE.			
RC	STRAIN-YT1;			
RX	MEDLINE=88225062; PubMed=3286255;			
RA	Kwon S.-T., Terada I., Matsuzawa H., Ohta T.;			
RT	"Nucleotide sequence of the gene for aqualysin I (a thermophilic alkaline serine protease) of Thermus aquaticus YT-1 and characteristics of the deduced primary structure of the enzyme.";			
RL	Eur. J. Biochem. 173:491-497(1988).			
RN	[3]			
RP	SEQUENCE OF 128-170.			
RX	MEDLINE=88151937; PubMed=3162211;			
RA	Matsuzawa H., Tokugawa K., Hamaoki M., Mizoguchi M., Taguchi H.,			
RT	Terada I., Kwon S.-T., Ohta T.;			
RT	"Purification and characterization of aqualysin I (a thermophilic alkaline serine protease) produced by Thermus aquaticus YT-1.";			
RL	Eur. J. Biochem. 171:441-447(1988).			
CC	-I- FUNCTION: AQUALYSIN I IS A THERMOPHILIC ALKALINE SERINE PROTEASE.			
CC	THE OPTIMAL TEMPERATURE FOR ITS CASEINOLYTIC ACTIVITY IS 80			
CC	DEGREES CELSIUS.			
CC	-I- SUBCELLULAR LOCATION: SECRETED.			
CC	-I- DEVELOPMENTAL STAGE: SECRETED FROM THE EARLY STATIONARY PHASE			
CC	UNTIL THE TIME THE CELLS CEASE TO GROW.			
CC	-I- PTM: THE N- AND C-TERMINAL PRO-SEQUENCES ARE REMOVED THROUGH THE			
CC	PROTEOLYTIC ACTIVITY OF AQUALYSIN I ITSELF, IN THAT ORDER. THE			
CC	C-TERMINAL PRO-SEQUENCE IS REQUIRED FOR TRANSLLOCATION OF THE			
CC	PROTEASES ACROSS THE OUTER MEMBRANE.			
CC	-I- PTM: TWO DISULFIDE BONDS ARE PRESENT.			
CC	-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE			
CC	SUBTILASE FAMILY.			

34	26	72.2	292	1	YTCZ_ECOLI	P39267 escherichia
35	26	72.2	298	1	IRGB_YIBCH	P25563 vibrio chol
36	26	72.2	302	1	SC14_CANGA	P53989 candida gla
37	26	72.2	367	1	GCST_MYCTU	Q10376 mycobacteri
38	26	72.2	368	1	UBPI_MOUSE	O9wv6 mus musculi
39	26	72.2	385	1	Y421_METTH	O26521 methanobact
40	26	72.2	400	1	FDH_PSESR	P33160 pseudomonas
41	26	72.2	429	1	DHE4_HALNI	O9hsn4 halobacteri
42	26	72.2	431	1	KRE2_CANAL	Q00310 candida alb
43	26	72.2	435	1	DHE4_HALSA	P29051 halobacteri
44	26	72.2	494	1	ADRO_MOUSE	O61578 mus musculi
45	26	72.2	505	1	MURC_SYNY3	P74528 synnechocyst
46	26	72.2	510	1	NIFB_FRAL	P46044 frankia aln
47	26	72.2	521	1	VENV_DHV11	P27427 horti virus
48	26	72.2	525	1	SERA_BACSU	P35136 bacillus su
49	26	72.2	562	1	TRZM_ERHME	Q47861 erwinia her
50	26	72.2	832	1	BGAL_ASPOP	P45582 asparagus o

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
CC EMBL; D90108; BAA14135.1; -
CC EMBL; X07734; CAA30559.1; -
CC PIR; S00620; S00620.
CC PIR; A35742; A35742.
CC HSP; P06873; 3PRK.
CC MEROPS; S08.051; -
CC InterPro; IPR000209; Peptidase_S8.
CC Pfam; PF00082; Peptidase_S8; 1.
CC PRINTS; PR00723; SUBTILISIN.
CC PROSITE; PS00136; SUBTILASE_ASP; 1.
CC PROSITE; PS00137; SUBTILASE_HIS; 1.
CC PROSITE; PS00138; SUBTILASE_SER; 1.
CC K0 Hydroxylase; Serine protease; zymogen; signal.
CC SIGNAL 1 14
CC FT PROPEP 15 127
CC FT CHAIN 128 408 AQUALYSIN I.
CC FT PROPEP 409 513
CC FT ACT_SITE 166 166 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 197 197 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 349 349 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC SEQUENCE 513 AA; 53913 MW; DDFDE6DA50B785 CRC64;

QY 1 wrxxaxyg 8
Db 495 WRIVAYSG 502

Query Match 86.1%; Score 31; DB 1; Length 513;
Best Local Similarity 62.5%; Pred. No. 9.1;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

RESULT 2
HMDH_PICUA STANDARD; PRT; 934 AA.
AC 074164;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE 3-HYDROXY-3-METHYLGUTARYL-COENZYME A REDUCTASE (EC 1.1.1.34) (HMG-COA
DE REDUCTASE).
GN HMG.
OS Pichia jadinii (Yeast) (Candida utilis).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OC NBL_TaxID=4903;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98320604; PubMed=9647847;
RA Shimada H., Kondo K., Fraser P.D., Miura Y., Saito T., Misawa N.;
RA "Increased carotenoid production by the food yeast *Candida utilis*
RT through metabolic engineering of the isoprenoid pathway."
RT Appl. Environ. Microbiol. 64:2676-2680(1998).
CC -1- FUNCTION: INVOLVED IN THE CONTROL OF CHOLESTEROL BIOSYNTHESIS. IT
CC IS THE RATE-LIMITING ENZYME OF THE STEROL BIOSYNTHESIS.
CC -1- CATALYTIC ACTIVITY: (R)-MEVALONATE + COA + 2 NADP(+) = (S)-3-
CC HYDROXY-3-METHYLGUTARYL-COA + 2 NADPH.
CC -1- PATHWAY: CHOLESTEROL BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
CC RETICULUM.
CC -1- SIMILARITY: BELONGS TO THE HMG-COA REDUCTASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

```

CC      EMBL: AB012603. BAA31937.1; -
CC      DR      InterPro: IPR002202; HMG-CoA_red.
CC      DR      InterPro: IPR000731; HMGCR_patched_5TM.
CC      DR      Pfam: PF00368; HMG-CoA_red. 1.
CC      DR      PRINTS: PR00071; HMGCOADPASE.
CC      DR      PROSITE: PS00066; HMG_COA_REDUCTASE_1; 1.
CC      DR      PROSITE: PS00318; HMG_COA_REDUCTASE_2; 1.
CC      DR      PROSITE: PS01192; HMG_COA_REDUCTASE_3; 1.
CC      DR      PROSITE: PS50065; HMG_COA_REDUCTASE_4; 1.
CC      DR      PROSITE: PS50156; SSD; 1.
CC      KM      Oxidoreductase; Glycoprotein; Endoplasmic reticulum; Transmembrane;
CC      KW      Cholesterol biosynthesis; NADP.
CC      FT      DOMAIN 1 442 MEMBRANE-BOUND.
CC      FT      DOMAIN 443 521 LINKER.
CC      FT      DOMAIN 522 934 CATALYTIC.
CC      FT      TRANSMEM 112 132 POTENTIAL.
CC      FT      TRANSMEM 142 162 POTENTIAL.
CC      FT      TRANSMEM 257 277 POTENTIAL.
CC      FT      TRANSMEM 335 355 POTENTIAL.
CC      FT      TRANSMEM 422 442 POTENTIAL.
CC      FT      ACT_SITE 618 618 BY SIMILARITY.
CC      FT      ACT_SITE 828 828 GENERAL BASE (BY SIMILARITY).
CC      FT      ACT_SITE 924 924 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      FT      CARBOHYD 364 364 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      FT      CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      FT      CARBOHYD 682 682 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      SQ      SEQUENCE 934 AA; 101233 MW; 3C3E43FC5623601C CRC64;

Query Match
Best Local Similarity 86.18; Score 31; DB 1; Length 934;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxayxg 8
Db 79 WRSRAYHG 86

RESULT 3
ID ZNUB_HAEIN STANDARD; PRT; 261 AA.
AC P44691;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 20-NOV-2001 (Rel. 40, Last annotation update)
DE HIGH-AFFINITY ZINC UPTAKE SYSTEM MEMBRANE PROTEIN ZNUB.
GN ZNUB OR H10407.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OC NCBI_TaxID=727;
OX [1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Klevavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu H.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Usterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Flie L.D., Fritchman J.L., Fuhrmann J.L., Georgagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus
RA influenzae Rd.";
```

```
RL Science 269:496-512(1995).
CC -1- FUNCTION: INVOLVED IN THE HIGH-AFFINITY ZINC UPTAKE TRANSPORT
CC SYSTEM (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE ABC-3 SUBFAMILY OF INTEGRAL MEMBRANE
CC PROTEINS. STRONG, TO E. COLI ZNOB.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32724; AAC22066.1; -.
DR TIGR: H10407; -.
DR InterPro: IPR001626; ABC-3.
DR Pfam: PF00950; ABC-3; 1.
DR Transport: 21nc; Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 8 28
FT TRANSMEM 54 74 POTENTIAL.
FT TRANSMEM 84 104 POTENTIAL.
FT TRANSMEM 125 145 POTENTIAL.
FT TRANSMEM 171 191 POTENTIAL.
FT TRANSMEM 214 234 POTENTIAL.
FT TRANSMEM 236 256 POTENTIAL.
SQ SEQUENCE 261 AA; 28465 MW; DD2AC8F0DF6CB96 CRC64;

Query Match 83.3%; Score 30; DB 1; Length 261;
Best Local Similarity 62.5%; Pred. No. 8;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
|||
DB 28 WRKWAYFG 35

RESULT 4
ID VTRAK_LAMB STANDARD; PRT; 199 AA.
AC P03729;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE TAIL ASSEMBLY PROTEIN K.
CN K.
OS Bacteriophage lambda.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;
CC Lambda phage group.
OX NCBI_TaxID-10710;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-83189071; PubMed-6221115;
RA Sanger F., Coulson A.R., Hong G.F., Petersen G.B.;
RT "Nucleotide sequence of bacteriophage lambda DNA.";
RL J. Mol. Biol. 163:729-773(1982).
CC -1- FUNCTION: GENE K PROTEIN IS INVOLVED IN THE ASSEMBLY OF THE
CC INITIATOR COMPLEX FOR TAIL POLYMERIZATION. IT HAS NOT BEEN FOUND
CC IN THE MATURE PHAGE.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-11 IS THE INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J02459; AAA96551.1; -.

```

```
DR PIR: A04355; TIBPKL.
DR InterPro: IPR000064; NLPC_P60.
DR Pfam: PF00877; NLPC_P60; 1.
SQ SEQUENCE 199 AA; 23011 MW; CEEB88F01E31ABAE CRC64;

Query Match 80.6%; Score 29; DB 1; Length 199;
Best Local Similarity 50.0%; Pred. No. 10;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
|||
DB 180 WRASAFYG 187

RESULT 5
ID CLD2_MOUSE STANDARD; PRT; 230 AA.
AC 088552;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-2.
GN CLDN2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID-10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-98311639; PubMed-9647647;
RA Furuse M., Fujita K., Hiragi T., Fujimoto K., Tsukita S.;
RT "Claudin-1 and -2: novel integral membrane proteins localizing at
RT tight junctions with no sequence similarity to occludin.";
RL J. Cell Biol. 141:1539-1550(1998).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF072128; AAC27079.1; -.
DR MGD: MGI:1276110; Cldn2.
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin; 1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN; 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 8 28
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 163 183 POTENTIAL.
SQ SEQUENCE 230 AA; 24483 MW; 38A7C074A1E0D5D2 CRC64;

Query Match 80.6%; Score 29; DB 1; Length 230;
Best Local Similarity 50.0%; Pred. No. 12;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
|||
DB 30 WRSSVYG 37

RESULT 6
AOPH_MYCTU

```

```

ID AOFH_MYCTU STANDARD: PRT; 454 AA.
AC O53320;
DT 15-DEC-1998 (Rel. 37, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE FLAVIN-CONTAINING MONOAMINE OXIDASE RV3170 (EC 1.4.3.-).
GN RV3170 OR MT3259 OR MTV014.14.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy J.,
RA Oliver S., Osborne K., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.B., Haft D., Hickey E.,
RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
CC -1- COFACTOR: FAD (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE FLAVIN MONOAMINE OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AL021646; CAAL1635.1; ALT_INIT.
DR EMBL; AE007139; AAK47598.1; -.
DR TIGR; MT3259; -.
DR TubercuList; RV3170; -.
DR InterPro; IPR002937; Amino_oxidase.
DR InterPro; IPR000205; NAD_binding.
DR Pfam; PF01593; Amino_oxidase; 1.
DR Hypothetical protein; oxidoreductase; Flavoprotein; FAD;
KW Complete proteome.
FT NP_BIND 14 69 FAD (ADP PART) (POTENTIAL).
FT SEQUENCE 454 AA; 49136 MW; 6C1AEB97FB2F435F CRC64;

```

```

Query Match 80.6%; Score 29; DB 1; Length 454;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

OY 1 WRXXAYXG 8
ID 312 WRASGYSG 319

```

```

RESULT 7
EX5B_ECOLI STANDARD: PRT; 1180 AA.
ID EX5B_ECOLI

```

```

AC P08394;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EXODEOXYRIBONUCLEASE V BETA CHAIN (EC 3.1.11.5) (EXODEOXYRIBONUCLEASE
DE V 135 KDA POLYPEPTIDE).
GN RECB OR RORR OR B2820.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87066729; PubMed=3537960;
RA Finch P.W., Storey A., Chapman K.E., Brown K., Hickson I.D.,
RA Emerson P.T.;
RT "Complete nucleotide sequence of the Escherichia coli recB gene.";
RL Nucleic Acids Res. 14:8573-8582(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=V1000;
RX MEDLINE=20229837; PubMed=10766864;
RA Arnold D.A., Kowalczykowski S.C.;
RT "Facilitated loading of RecA protein is essential to recombination by
RT RecBCD enzyme.";
RL J. Biol. Chem. 275:12261-12265(2000).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE OF 1-11 FROM N.A.
RX MEDLINE=87040734; PubMed=3534791;
RA Finch P.W., Wilson R.E., Brown K., Hickson I.D., Emerson P.T.;
RT "Complete nucleotide sequence of the Escherichia coli ptr gene
RT encoding protease III.";
RL Nucleic Acids Res. 14:7695-7703(1986).
RN [5]
RP SEQUENCE OF 1093-1180 FROM N.A.
RX MEDLINE=87066730; PubMed=3537961;
RA Finch P.W., Storey A., Brown K., Hickson I.D., Emerson P.T.;
RT "Complete nucleotide sequence of recD, the structural gene for the
RT alpha subunit of Exonuclease V of Escherichia coli.";
RL Nucleic Acids Res. 14:8583-8594(1986).
CC -1- FUNCTION: REQUIRED FOR EFFICIENT DNA REPAIR. IT CATALYZES THE
CC UNWINDING OF DOUBLE-STRANDED DNA AND THE CLEAVAGE OF SINGLE-
CC STRANDED DNA AND IT STIMULATES LOCAL GENETIC RECOMBINATION.
CC ALL OF THESE ACTIVITIES REQUIRE CONCOMITANT HYDROLYSIS OF ATP.
CC -1- CATALYTIC ACTIVITY: EXONUCLEOLYTIC CLEAVAGE (IN THE PRESENCE OF
CC ATP) IN EITHER 5'-TO 3'-OR 3'-TO 5'-DIRECTION TO YIELD 5'-
CC PHOSPHOLIGONUCLEOTIDES.
CC -1- SUBUNIT: CONSIST OF THREE SUBUNITS: RECB, RECC AND RECD.
CC -1- SIMILARITY: BELONGS TO THE UVRD SUBFAMILY OF HELICASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial/
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X04581; CAA28250.1; -.
DR EMBL; AF179304; AAD56369.1; -.
DR EMBL; U29581; AAB40467.1; -.
DR EMBL; AE000365; AAC75859.1; -.
DR EMBL; X06227; CAA29577.1; -.

```



```
DR EMBL: X04582; CAA28252.1; -.
DR PIR: A25532; NCECKS.
DR HSSP: P56255; 1PJR.
DR Ecogen: BC10824; recB.
DR InterPro: IPR000212; UVRD-helicase.
DR Pfam: PF00580; UVRD-helicase; 1
KW Hydrolase; Nuclease; Exonuclease; Endonuclease; Helicase; ATP-binding;
KW DNA repair; Complete proteome.
FT NP_BIND 23
FT SEQUENCE 1180 AA; 133958 MW; F9AC331808EF281 CRC64;
SO
OY 1 wrxxayxg 8
DB 901 WRVTSYSG 908
OY 1 wrxxayxg 8
DB 901 WRVTSYSG 908
RESULT 8
ID VA28_VARY STANDARD; PRT; 146 AA.
AC P33847;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE PROTEIN A28.
GN A28L OR A31L OR A31.5L.
OS Variola virus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10255;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-INDIA-1967 / ISOLATE IND3;
RA MEDLINE=92209372; PubMed=166548;
RA Shchelkunov S.N., Marennikova S.S., Tolmenin A.V., Bilnov V.M.,
RA Chelikhov V.E., Gutov V.V., Safonov P.F., Pozdnyakov S.G.,
RA Shchelkha E.M., Gashnikov P.V., Anaparidze O.G., Sandakhchlev L.S.;
RT "Creation of a clone library of fragments from the natural variola
RT virus and study of the structural and functional organization of
RT viral genes from a circle of hosts.";
RL Dokl. Akad. Nauk SSSR 321:402-406(1991).
RN [2]
RP COMPLETE GENOME.
RC STRAIN-INDIA-1967 / ISOLATE IND3;
RA MEDLINE=93202281; PubMed=8384129;
RA Shchelkunov S.N., Bilnov V.M., Sandakhchlev L.S.;
RT "Genes of variola and vaccinia viruses necessary to overcome the host
RT protective mechanisms.";
RL FEBS Lett. 319:80-83(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-HARVEY;
RA MEDLINE=93057361; PubMed=1331292;
RA Aguado B., Selmes I.P., Smith G.L.;
RT "Nucleotide sequence of 21.8 kbp of variola major virus strain Harvey
RT and comparison with vaccinia virus.";
RL J. Gen. Virol. 73:2887-2902(1992).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN BANGLADESH-1975;
RA MEDLINE=94088747; PubMed=8264798;
RA Masung R.F., Esposito J.J., Liu L., Qi J., Uterback T.R.,
RA Knight J.C., Aubin L., Yuran T.E., Parsons J.M., Loparev V.N.,
RA Sellanov N.A., Cavallari K.F., Kertavage A.R., Mahy B.W.J.,
RA Venter C.J.;
RT "Potential virulence determinants in terminal regions of variola
RT smallpox virus genome.";
RL Nature 366:748-751(1993).
RN [5]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN-GARCIA-1966;
RA Shchelkunov S.N., Tolmenin A.V., Resenchuk S.M., Bilnov V.M.,
RA Sandakhchlev L.S.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO CAPRIPPOXVIRUS (STRAIN KS-1) PROTEIN HM3 AND TO
CC AMSACTA MOOREI ENTOMOPOXVIRUS PROTEIN GAR.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X69198; CAA49076.1; -.
DR EMBL: X67115; CAA47502.1; -.
DR EMBL: L22579; AAA60883.1; -.
DR EMBL: X76266; CAA53857.1; -.
DR PIR: E36851; E36851.
DR PIR: J01834; J01834.
DR PIR: S46858; S46858.
SO SEQUENCE 146 AA; 16244 MW; B8121A3A5D75A365 CRC64;
OY 1 wrxxayxg 8
DB 73 WRVAVPG 80
OY 1 wrxxayxg 8
DB 73 WRVAVPG 80
RESULT 9
ID CLD1_HUMAN STANDARD; PRT; 211 AA.
AC O95832;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CLAUDIN-1 (SENESCENCE-ASSOCIATED EPITHELIAL MEMBRANE PROTEIN).
GN CLDN1 OR CLD1 OR SEMP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=9913301; PubMed=9931503;
RA Swisshelm K.L., Machl A., Planitzner S., Robertson R., Kubies M.,
RA Hoeler S.;
RT "SEMP1, a senescence-associated cDNA isolated from human mammary
RT epithelial cells, is a member of an epithelial membrane protein
RT superfamily.";
RL Gene 226:285-295(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC "Mitic L.M., Anderson J.M.";
RT "Human claudin-1 isolated from Caco-2 mRNA.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC MEDLINE=20290992; PubMed=10828592;
RA Halford S., Spencer P., Greenwood J., Winton H., Hunt D.M.,
RA Adamson P.;
RT "Assignment(1) of claudin-1 (CLDN1) to human chromosome 3q28-->q29
RT with somatic cell hybrids.";
RL Cytogenet. Cell Genet. 88:217-217(2000).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF101051; AAD16433.1; -
CC EMBL: AF115546; AAD22962.1; -
CC EMBL: AF134160; AAF61393.1; -
CC MIM: 603718; -
CC InterPro: IPR001832; Claudin.
CC InterPro: IPR000729; PMP22_Claudin.
CC Pfam: PF00822; PMP22_Claudin; 1.
CC PRINTS: PR01077; CLAUDIN.
CC PROSITE: PS01346; CLAUDIN; 1.
CC TIGHT junction; Transmembrane.
CC TRANSMEM 8 28 POTENTIAL.
CC TRANSMEM 82 102 POTENTIAL.
CC TRANSMEM 116 136 POTENTIAL.
CC TRANSMEM 164 184 POTENTIAL.
CC CONFLICT 62 62 I -> V (IN REF. 2).
CC CONFLICT 135 135 V -> A (IN REF. 2).
CC SEQUENCE 211 AA; 22744 MW; 0726900DB6C214F0 CRC64;
QY 1 wrxxayxg 8
| | | |
| | | |
Db 30 WRXSYAG 37
Query Match 77.8%; Score 28; DB 1; Length 211;
Best Local Similarity 50.08; Pred. No. 18;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
RESULT 10
CLDX_BRAE STANDARD; PRT; 215 AA.
AC OGYH92;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CLAUDIN-LIKE PROTEIN ZFA422 (CLAUDIN 7).
OS Brachydanio rerio (zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RA SEQUENCE FROM N.A.
RA Keen T.J., Inglehart C.F.;
RA Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.
RL [2]
RE SEQUENCE FROM N.A.
RP MEDLINE=20525589; PubMed=11071763;
RX Chin A.J., Tsang M., Weinberg E.S.;
RT "Heart and gut chiralities are controlled independently from initial
RT heart position in the developing zebrafish.";
RL Dev. Biol. 227:403-421(2000).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: AJ011788; CAA09776.1; -
DR EMBL: AE260240; AAG24512.1; -
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin; 1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN; 1.
CC TIGHT junction; Transmembrane.
CC TRANSMEM 8 28 POTENTIAL.
CC TRANSMEM 82 102 POTENTIAL.
CC TRANSMEM 118 138 POTENTIAL.
CC TRANSMEM 163 183 POTENTIAL.
CC SEQUENCE 215 AA; 22865 MW; BC04870B75B8CB9D CRC64;
QY 1 wrxxayxg 8
| | | |
| | | |
Db 30 WRXSAVVG 37
Query Match 77.8%; Score 28; DB 1; Length 215;
Best Local Similarity 50.08; Pred. No. 19;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
RESULT 11
YVD3_CAEEL STANDARD; PRT; 551 AA.
ID YVD3_CAEEL
AC P55114;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL ZINC METALLOPROTEINASE K04E7.3 PRECURSOR (EC 3.4.24.-).
GN K04E7.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodera; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RA SEQUENCE FROM N.A.
RA STRAIN-BRISTOL N2;
RL Nham M.;
RL Submitted (Oct-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12A (ZINC
CC METALLOPROTEASE); ALSO KNOWN AS THE ASTACIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U39666; AAB80412.1; -
CC HSSP: P00740; IIXA.
DR WORMPEP: K04E7.3; CE02798.
DR InterPro: IPR001506; Astacin.
DR InterPro: IPR000859; CUB.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000884; TSP1.
DR InterPro: IPR000130; Zh_mtpeptdse.
CC Pfam: PF01400; Astacin; 1.
CC Pfam: PF00431; CUB; 1.
DR SMART: SM00042; CUB; 1.
DR SMART: SM00209; TSP1; 1.
DR SMART: SM00235; ZMNC; 1.
DR PROSITE: PS01180; CUB; FALSE_NEG.
DR PROSITE: PS00022; EGF_1; 1.
DR PROSITE: PS01186; EGF_2; FALSE_NEG.
DR PROSITE: PS50092; TSP1; 1.

DR PROSITE: PS00142: ZINC_PROTEASE. 1.
KW Hypothetical protein; Hydrolase; Metalloprotease; Zinc;
KW Signal; EGF-like domain; Glycoprotein.
FT SIGNAL 1 ?
FT CHAIN ? 551
FT ?
FT DOMAIN 18 28 KOA07.3.
FT DOMAIN 318 358 POLY-PRO.
FT DOMAIN 359 484 EGF-LIKE.
FT DOMAIN 485 535 CUB.
FT METAL 219 219 TSP TYPE-1.
FT ACT_SITE 220 220 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 223 223 BY SIMILARITY.
FT METAL 229 229 ZINC (CATALYTIC) (BY SIMILARITY).
FT CARBOHYD 136 136 ZINC (CATALYTIC) (BY SIMILARITY).
FT CARBOHYD 378 378 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 378 378 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 551 AA: 61673 MW: 70C9A16B56C8B7E CRC64;

Query Match 77.8%; Score 28; DB 1; Length 551;
Best Local Similarity 50.0%; Pred. No. 46;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxyxyg 8
ID 1 1 1
DB 376 WRN1SYSG 383

RESULT 12
XFP_LACLA STANDARD; PRT; 822 AA.

AC 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE XYULOSE-5-PHOSPHATE/FRUCTOSE-6-PHOSPHATE PHOSPHOKETOLASE
(EC 4.1.2.9) (EC 4.1.2.22).
XP OR PTK.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1360;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=IL1403;
RC MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Manger S., Jallion O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RA "The complete genome sequence of the lactic acid bacterium Lactococcus
lactis ssp. lactis IL1403.";
RL Genome Res. 11:731-753(2001).
CC -1- FUNCTION: PHOSPHOKETOLASE USING BOTH FRUCTOSE 6-PHOSPHATE AND
XYULOSE 5-PHOSPHATE AS SUBSTRATE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: D-XYULOSE 5-PHOSPHATE + PHOSPHATE -> ACETYL
PHOSPHATE + D-GLYCERALDEHYDE 3-PHOSPHATE + H(2)O.
CC -1- CATALYTIC ACTIVITY: D-FRUCTOSE 6-PHOSPHATE + PHOSPHATE -> ACETYL
PHOSPHATE + D-ERYTHROSE 4-PHOSPHATE + H(2)O.
CC -1- COFACTOR: THIAMINE PYROPHOSPHATE (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE XFP FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: AE006381; AAK05600.1; -
DR PROSITE: PS00187; TPP-ENZYMES; FALSE_NGC.
KW Lyase; Flavoprotein; Thiamine pyrophosphate; Complete proteome.
SO SEQUENCE 822 AA: 93363 MW: C686D565D3E8E22F CRC64;

Query Match 77.8%; Score 28; DB 1; Length 822;
Best Local Similarity 50.0%; Pred. No. 68;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wxxyxyg 8
ID 1 1 1
DB 17 WRATYLG 24

RESULT 13
HIRA_CHICK STANDARD; PRT; 1018 AA.
ID HIRA_CHICK
AC P79987;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 38, Last annotation update)
DE HIRA PROTEIN (TPP1 LIKE ENHANCER OF SPLIT PROTEIN 1) (CHINA).
GN HIRA OR TUPPEL.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=97217783; PubMed=9063744;
RA Roberts C., Daw S.C., Halford S., Scambler P.J.;
RT "Cloning and developmental expression analysis of chick Hira (Chlra),
RT a candidate gene for DiGeorge syndrome.";
RL Hum. Mol. Genet. 6:237-245(1997).
CC -1- FUNCTION: COULD PLAY A PART IN MECHANISMS OF TRANSCRIPTIONAL
CC REGULATION SIMILAR TO THAT PLAYED BY YEAST HIR1 AND HIR2 TOGETHER.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: CONTAINS 4 WD REPEATS (TPP-ASP DOMAINS).
CC -1- SIMILARITY: BELONGS TO THE HIR1 FAMILY OF WD-REPEAT PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: X99375; CAA67754.1; -
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40.7.
DR PRINTS: PR00320; GPROTEINBRPT.
DR SMART: SM00320; WD40.6.
DR PROSITE: PS00678; WD_REPEATS.1; 1.
DR PROSITE: PS50082; WD_REPEATS.2; 3.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
KW Transcription regulation; Repeat; WD repeat; Nuclear protein.
FT REPEAT 68 98 WD 1.
FT REPEAT 128 158 WD 2.
FT REPEAT 172 202 WD 3.
FT REPEAT 266 313 WD 4.
FT DOMAIN 267 286
FT DOMAIN 634 651 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 407 413 POLY-GLN.
FT DOMAIN 552 555 POLY-SER.
FT DOMAIN 644 647 POLY-LYS.
SO SEQUENCE 1018 AA: 11817 MW: 58C6C710A8FA7BF1 CRC64;

Query Match 77.8%; Score 28; DB 1; Length 1018;
Best Local Similarity 50.0%; Pred. No. 83;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxyxyg 8
ID 1 1 1
DB 97 WKRAAYIG 104

```

RESULT 14
VAT_CAMVP STANDARD; PRT: 159 AA.
ID VAT_CAMVP
AC P19818;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE APHID TRANSMISSION PROTEIN (PROTEIN 2).
OS II.
GN Cauliflower mosaic virus (strain PV47) (CaMV).
OC Viruses; Retroid viruses; Caulimovirus.
OX NCBI_TaxID=10647;
RN [1]
RA SEQUENCE FROM N.A.
RP Modjatedi N., Volovitch M., Mazolini L., Yot P.;
RT "Comparison of the predicted secondary structure of aphid
RT transmission factor for transmissible and non-transmissible
RT cauliflower mosaic virus strains.";
RL FEBS Lett. 181:223-228(1985).
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN VIRUS TRANSMISSION.
CC -!- SIMILARITY: BELONGS TO THE CAULIMOVIRUSES ORF II FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: M37581; AAA9697.1; -
SO SEQUENCE 159 AA; 17843 MW; 196DF6D1F9D0B15A CRC64;

Query Match 75.0%; Score 27; DB 1; Length 159;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
| | | | |
DB 61 WRINSTFG 68

RESULT 15
COX2_SUTLAC STANDARD; PRT: 168 AA.
ID COX2_SUTLAC
AC P39479;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE QUINOL OXIDASE POLYPEPTIDE II (EC 1.9.3.-) (CYTOCHROME AA3 SUBUNIT 2)
DE (OXIDASE AA(3) SUBUNIT 2).
OS SOXA.
GN Sulfolobus acidocaldarius.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=2285;
RN [1]
RA SEQUENCE FROM N.A.
RP STRAIN-ATCC 33909 / NCIB 11770 / DSM 639;
RX MEDLINE=92192013; PubMed=1372250;
RA Luebben M., Kolmerer B., Saraste M.;
RT "An archaeobacterial terminal oxidase combines core structures of two
RT mitochondrial respiratory complexes.";
EMBO J. 11:805-812(1992).
CC -!- FUNCTION: THE TERMINAL OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX.
CC -!- FUNCTION: SUBUNIT 2 TRANSFERS THE ELECTRONS FROM CALDARIELLA
CC QUINOL TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1 THAT IS
CC FORMED BY HEME A3 AND CU(B).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

```

```

CC -!- SIMILARITY: SOME, TO MITOCHONDRIAL OR BACTERIAL COX2 SUBUNITS.
CC BUT LACK HEME-BINDING DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: X62643; CAA44509.1; -
DR PIR: S21041; S21041.
KW Oxidoreductase; Transmembrane; Respiratory chain; Electron transport.
FT TRANSMEM 9
SQ SEQUENCE 168 AA; 18874 MW; 7B5C98CA24D06846 CRC64;

Query Match 75.0%; Score 27; DB 1; Length 168;
Best Local Similarity 50.0%; Pred. No. 25;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
| | | | |
DB 147 WRDAEYAG 154

RESULT 16
CLD4_CERAE STANDARD; PRT: 209 AA.
ID CLD4_CERAE
AC 019005;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-4 (CLOSTRIDIUM PERFRINGENS ENTEROTOXIN RECEPTOR) (CPE-
DE RECEPTOR) (CPE-R).
GN CLDN4 OR CPEPR1 OR CPER.
OS Cercarial dermatitis (Swimmer's itch) (Green monkey) (Grivel).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Cercopithecus.
OX NCBI_TaxID=9534;
RN [1]
RA SEQUENCE FROM N.A.
RX MEDLINE=97242441; PubMed=9087440;
RA Katohira J., Inoue N., Horikuchi Y., Matsuda M., Sugimoto N.;
RT "Molecular cloning and functional characterization of the receptor for
RT Clostridium perfringens enterotoxin.";
RL J. Cell Biol. 136:1239-1247(1997).
CC -!- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: D88492; BAA22781.1; -
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin; 1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN; 1.
CC Tight junction; Transmembrane.
FT TRANSMEM 10
FT TRANSMEM 82
FT TRANSMEM 118
FT TRANSMEM 161
SQ SEQUENCE 209 AA; 22029 MW; 474DB309F95289E CRC64;

```

```
Query Match      75.0%: Score 27; DB 1; Length 209;
Best Local Similarity 50.0%: Pred. No. 30;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
   || | | |
DB 30 WRVTAFIG 37

RESULT 17
CLD4_HUMAN
ID CLD4_HUMAN STANDARD: PRT; 209 AA.
AC O14493;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-4 (CLOSTRIDIUM PERFRINGENS ENTEROTOXIN RECEPTOR) (CPE-
DE RECEPTOR) (CPE-R).
GN CLDN4 OR CPETR1 OR CPER.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
CC [1]
CC SEQUENCE FROM N.A.
CC TISSUE=Fetal brain;
CC MEDLINE=97476271; PubMed=9334247;
RA Katsuhira J., Sugiyama H., Inoue N., Horiguchi Y., Matsuda M.,
RA Sugimoto N.;
RT "Clostridium perfringens enterotoxin utilizes two structurally related
RT membrane proteins as functional receptors in vivo.";
RL J. Biol. Chem. 272:26552-26558(1997).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AB000712; BAA22984.1; -.
CC MIM: 602909; -.
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin; 1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN; 1.
DR PROSITE: PS01346; CLAUDIN; 1.
KM Tight junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 118 138 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
SO SEQUENCE 209 AA; 22077 MW; 0659A93AA5F0E4C5 CRC64;
```

```
AC O9YH90;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-LIKE PROTEIN ZF-A9.
DE CLAUDIN-LIKE PROTEIN ZF-A9.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
CC [1]
CC SEQUENCE FROM N.A.
CC Keen T.J., Inglehearn C.F.;
CC Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ011790; CAA09778.1; -.
CC InterPro: IPR001832; Claudin.
CC InterPro: IPR000729; PMP22_Claudin.
CC Pfam: PF00822; PMP22_Claudin; 1.
CC PRINTS: PR01077; CLAUDIN.
CC PROSITE: PS01346; CLAUDIN; 1.
KM Tight junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 81 101 POTENTIAL.
FT TRANSMEM 114 134 POTENTIAL.
FT TRANSMEM 159 179 POTENTIAL.
SO SEQUENCE 209 AA; 22091 MW; C413143811853D58 CRC64;
```

```
Query Match      75.0%: Score 27; DB 1; Length 209;
Best Local Similarity 50.0%: Pred. No. 30;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
   || | | |
DB 29 WRVTAFIG 36

RESULT 19
CLD4_MOUSE
ID CLD4_MOUSE STANDARD: PRT; 210 AA.
AC O35054;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-4 (CLOSTRIDIUM PERFRINGENS ENTEROTOXIN RECEPTOR) (CPE-
DE RECEPTOR) (CPE-R).
GN CLDN4 OR CPETR1 OR CPER.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
CC [1]
CC SEQUENCE FROM N.A.
CC MEDLINE=97476271; PubMed=9334247;
RA Katsuhira J., Sugiyama H., Inoue N., Horiguchi Y., Matsuda M.,
RA Sugimoto N.;
RT "Clostridium perfringens enterotoxin utilizes two structurally related
RT membrane proteins as functional receptors in vivo.";
RL J. Biol. Chem. 272:26552-26558(1997).
CC [2]
CC SEQUENCE FROM N.A.
```

RA MEDLINE=99110921; PubMed=9892664;
RA Morita K., Furuse M., Fujimoto K., Tsukita S.;
RT "Claudin multigene family encoding four-transmembrane domain protein
CC components of tight junction strands";
CC Proc. Natl. Acad. Sci. U.S.A. 96:511-516(1999).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB000713; BAA22985.1; -;
DR EMBL; AF087822; AAD09757.1; -;
DR MGD; MGI:131314; Cldn4.
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam; PF00822; PMP22_Claudin; 1.
DR PRINTS; PR01077; CLAUDIN.
DR PROSITE; PS01346; CLAUDIN; 1.
DR Tight junction; Transmembrane.
KW TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
SQ SEQUENCE 210 AA; 22338 MW; 3B6D571BC71D6564 CRC64;

Query Match 75.0%; Score 27; DB 1; Length 210;
Best Local Similarity 50.0%; Pred. No. 31;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 wxxxayxg 8
| | | | |
Db 30 WRVTAFIG 37

RESULT 20
CLD3_MOUSE
ID CLD3_MOUSE STANDARD; PRT; 219 AA.
AC 0920G9;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-3 (CLOSTRIDIUM PERFRINGENS ENTEROTOXIN RECEPTOR 2) (CPE-
DE RECEPTOR 2) (CPE-R 2).
GN CLDN3 OR CPETR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99097345; PubMed=9878248;
RA Paperna T., Peoples R., Wang Y.K., Kaplan P., Francke U.;
RT "Genes for the CPE receptor (CPETR1) and the human homolog of RVP1
RT (CPETR2) are localized within the Williams-Beuren syndrome deletion.";
RL Genomics 54:453-459(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99110921; PubMed=9892664;
RA Morita K., Furuse M., Fujimoto K., Tsukita S.;
RT "Claudin multigene family encoding four-transmembrane domain protein
RT components of tight junction strands";
RT Proc. Natl. Acad. Sci. U.S.A. 96:511-516(1999).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF095905; AAD14608.1; -;
DR EMBL; AF087821; AAD09756.1; -;
DR MGD; MGI:1329044; Cldn3.
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam; PF00822; PMP22_Claudin; 1.
DR PRINTS; PR01077; CLAUDIN.
DR PROSITE; PS01346; CLAUDIN; 1.
DR Tight junction; Transmembrane.
KW TRANSMEM 9 29 POTENTIAL.
FT TRANSMEM 81 101 POTENTIAL.
FT TRANSMEM 116 136 POTENTIAL.
FT TRANSMEM 160 180 POTENTIAL.
SQ SEQUENCE 219 AA; 22284 MW; 62F67810D9B9BD37 CRC64;

Query Match 75.0%; Score 27; DB 1; Length 219;
Best Local Similarity 50.0%; Pred. No. 32;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 wxxxayxg 8
| | | | |
Db 29 WRVSAFTG 36

Search completed: January 14, 2002, 07:40:40
Job time: 506 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: January 14, 2002, 07:39:46 ; Search time 112.89 Seconds
(without alignments)
10.366 Million cell updates/sec

Title: 09-185908-1f
Perfect score: 36
Sequence: 1 wrxxayxg 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database :
1: SP archaea: *
2: SP bacteria: *
3: SP fungi: *
4: SP human: *
5: SP invertebrate: *
6: SP mammal: *
7: SP mhc: *
8: SP organelle: *
9: SP phage: *
10: SP plant: *
11: SP rodent: *
12: SP virus: *
13: SP vertebrate: *
14: SP unclassified: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	33	91.7	613 2 09A956	09A956 caulobacter
2	32	88.9	479 12 09Q0N6	09Q0N6 sugarcane y
3	32	88.9	479 12 09JH75	09JH75 sugarcane y
4	32	88.9	497 4 09BAR6	09BAR6 homo sapien
5	30	83.3	260 2 09K0R7	09K0R7 vibrio chol
6	30	83.3	261 2 09CP25	09CP25 pasteurilla
7	30	83.3	262 2 09HT72	09HT72 pseudomonas
8	30	83.3	303 10 064761	064761 arabidopsis
9	30	83.3	660 2 09RD13	09RD13 streptomyc
10	30	83.3	938 10 09LDE7	09LDE7 oryza sativ
11	30	83.3	955 10 09XE23	09XE23 oryza sativ
12	30	83.3	1011 10 09FRA2	09FRA2 oryza sativ
13	30	83.3	1281 10 09LDA3	09LDA3 oryza sativ
14	30	83.3	1591 10 09LDW9	09LDW9 oryza sativ
15	30	83.3	1626 10 09FW81	09FW81 oryza sativ
16	30	83.3	1641 10 09AYG3	09AYG3 oryza sativ
17	29	80.6	65 2 09RD02	09RD02 streptomyc
18	29	80.6	65 2 09RD18	09RD18 streptomyc
19	29	80.6	65 2 09RD17	09RD17 streptomyc

20	29	80.6	1208 2 09KPP6	09KPP6 vibrio chol
21	28	77.8	64 7 030824	030824 ovis aries
22	28	77.8	266 2 09EWM0	09EWM0 streptomyc
23	28	77.8	278 1 058458	058458 pyrococcus
24	28	77.8	303 12 084711	084711 pea enation
25	28	77.8	366 2 09AAC7	09AAC7 caulobacter
26	28	77.8	431 2 09F9H3	09F9H3 burkholderi
27	28	77.8	493 12 003708	003708 pea enation
28	28	77.8	493 12 093184	093184 pea enation
29	28	77.8	648 10 049559	049559 arabidopsis
30	28	77.8	822 2 09CFH4	09CFH4 lactococcus
31	27	75.0	64 7 030825	030825 ovis aries
32	27	75.0	91 7 019588	019588 homo sapien
33	27	75.0	116 2 09RPY2	09RPY2 brucella su
34	27	75.0	154 2 09K199	09K199 neisseria m
35	27	75.0	154 2 09JSM8	09JSM8 neisseria m
36	27	75.0	158 2 09CCU7	09CCU7 mycobacteri
37	27	75.0	159 9 038144	038144 bacterioph
38	27	75.0	165 7 09MW45	09MW45 homo sapien
39	27	75.0	166 11 099RK2	099RK2 mus musculu
40	27	75.0	169 2 007698	007698 mycobacteri
41	27	75.0	175 8 079714	079714 rhea americ
42	27	75.0	192 2 066846	066846 aquifex aeo
43	27	75.0	212 7 095733	095733 homo sapien
44	27	75.0	213 5 09NAQ9	09NAQ9 caenorhabdl
45	27	75.0	213 7 019633	019633 homo sapien
46	27	75.0	214 13 09DE12	09DE12 xenopus lae
47	27	75.0	214 13 098SR2	098SR2 gallus gall
48	27	75.0	246 2 09S1N7	09S1N7 streptomyc
49	27	75.0	274 10 09SW79	09SW79 trillium ae
50	27	75.0	330 2 09S2V4	09S2V4 streptomyc

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	613 AA.
ID	09A956			
AC	09A956;			
DT	01-JUN-2001 (TREMREL. 17, Created)			
DT	01-JUN-2001 (TREMREL. 17, Last sequence update)			
DT	01-JUN-2001 (TREMREL. 17, Last annotation update)			
DE	TONB-DEPENDENT RECEPTOR, PUTATIVE.			
GN	CC1138.			
OS	Caulobacter crescentus.			
OC	Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;			
OC	Caulobacter			
OX	NCBI_TaxID=69394;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21173698; PubMed=11259647;			
RA	Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,			
RA	Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,			
RA	Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,			
RA	DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,			
RA	Kolony J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,			
RA	Utecherback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,			
RA	Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;			
KT	*Complete genome sequence of Caulobacter crescentus.*;			
RL	Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).			
DR	EMBL; AE005792; AAK23122.1; -.			
DR	TIGR; CC1138; -.			
KW	Receptor; Complete proteome.			
SQ	SEQUENCE 613 AA; 64411 MW; ED65208A3D82B97D CRC64;			

Query Match 91.7%; Score 33; DB 2; Length 613;
Best local Similarity 62.5%; Pred. No. 27;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 wrxxayxg 8
|| || |

Db 370 WRAAYAG 377

RESULT 2

ID Q9QON6 PRELIMINARY; PRT; 479 AA.

AC Q9QON6;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DE PUTATIVE APHID TRANSMISSION FACTOR (FRAGMENT).

OS sugarcane yellow leaf virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;

OC Unassigned Luteoviridae.

OX NCBI_TaxID=94290;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-A;

RA Moonan F., Molina J.J., Mirov T.E.;

RT "Sugarcane yellow leaf virus is a new virus with a genome that has Poliovirus, Luteovirus, and Enamovirus properties."

RT Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF157029; AAD45687.1; -

DR InterPro: IPR000893; Luteo_ORF6.

DR InterPro: IPR002929; PLRV_ORF5.

DR Pfam: PF01690; PLRV_ORF5; 1.

DR PRINTS: PR00910; LVIRUSORF6.

DR PRINTS: PR01217; PRICHEXTENSN.

FT NON_TER 1

SO SEQUENCE 479 AA; 51676 MW; 6C1DBAAA2BA5EEC9 CRC64;

Query Match 88.9%; Score 32; DB 12; Length 479;

Best Local Similarity 62.5%; Pred. No. 35;

Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8

Db 129 WRAVAYSG 136

RESULT 3

ID Q9JH75 PRELIMINARY; PRT; 479 AA.

AC Q9JH75;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DE PUTATIVE APHID TRANSMISSION FACTOR (FRAGMENT).

OS sugarcane yellow leaf virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;

OC Unassigned Luteoviridae.

OX NCBI_TaxID=94290;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-CV. CP65-357;

RA Smith G.R.;

RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-CV. CP65-357;

RC MEDLINE=20318675; PubMed=10859394;

RA Smith G.R., Borg Z., Lockhart B.E.L., Brathwaite K.S., Gibbs M.J.;

RT "Sugarcane yellow leaf virus: a novel member of the Luteoviridae that probably arose by inter-species recombination.";

RT J. Gen. Virol. 81:1865-1869(2000).

DR EMBL: AJ249447; CAB75437.1; -

DR InterPro: IPR000893; Luteo_ORF6.

DR InterPro: IPR002929; PLRV_ORF5.

DR InterPro: IPR002965; P_Rich_extensn.

DR Pfam: PF01690; PLRV_ORF5; 1.

DR PRINTS: PR00910; LVIRUSORF6.

DR PRINTS: PR01217; PRICHEXTENSN.

FT NON_TER 1

SO SEQUENCE 479 AA; 51616 MW; AD3B98A1B658652F CRC64;

Query Match 88.9%; Score 32; DB 12; Length 479;

Best Local Similarity 62.5%; Pred. No. 35;

Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8

Db 129 WRAVAYSG 136

RESULT 4

ID Q9BRR6 PRELIMINARY; PRT; 497 AA.

AC Q9BRR6;

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DE SIMILAR TO RIKEN CDNA 261001/G09 GENE.

OS Homo sapiens (Human).

OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=KIDNEY ADENOCARCINOMA;

RA Strausberg R.;

RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: BC006112; AAH06112.1; -

SO SEQUENCE 497 AA; 54088 MW; B758E977CDA88F8F CRC64;

Query Match 88.9%; Score 32; DB 4; Length 497;

Best Local Similarity 62.5%; Pred. No. 37;

Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8

Db 4 WRGSAYAG 11

RESULT 5

ID Q9KOB7 PRELIMINARY; PRT; 260 AA.

AC Q9KOB7;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DE ZINC ABC TRANSPORTER, PERMEASE PROTEIN.

GN VC2083.

OS Vibrio cholerae.

OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.

OX NCBI_TaxID=666;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=EL TOR N16961 / SEROTYPE O1;

RX MEDLINE=20406833; PubMed=10952301;

RA Heidelberg J.F., Eissen J.A., Nelson W.C., Clayton R.A., Gwin M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A., Gill S.R., Nelson K.E., Read T.D., Tetteilin H., Richardson D., Ernoolaeva M.D., Yamachavan J., Bass S., Qin H., Dragoi I., Sellers P., McDonald L., Utleback T., Fleischmann R.D., Niernan W.C., White O., Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C., Fraser G.M.;

RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae";

RL Nature 406:477-483(2000).

DR EMBL: AE004282; AAF95229.1; -

DR TIGR: VC2083; -

DR InterPro: IPR001626; ABC-3.

DR Pfam: PF00950: ABC-3; 1.
KW Complete proteome.
SQ SEQUENCE 260 AA; 27718 MW; 951B5C185C235957 CRC64;

Query Match
Best Local Similarity 83.3%; Score 30; DB 2; Length 260;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxaxyxg 8
|| || |
Db 28 WRRMAYFG 35

RESULT 6
ID 09CP25 PRELIMINARY; PRT; 261 AA.
AC 09CP25;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DE 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE HYPOTHETICAL PROTEIN PM0241.
GN PM0241.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.-J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL: AE006058; AAK0325.1; -
DR InterPro: IPR001626; ABC-3.
DR Pfam: PF00950; ABC-3; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 261 AA; 28440 MW; 25309B50B39A7223 CRC64;

Query Match
Best Local Similarity 83.3%; Score 30; DB 2; Length 261;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxaxyxg 8
|| || |
Db 28 WRRMAYFG 35

RESULT 7
ID 09HT72 PRELIMINARY; PRT; 262 AA.
AC 09HT72;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE PERMEASE OF ABC ZINC TRANSPORTER ZNUB.
GN ZNUB OR PA5501.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
Hickey M.J., Brinkman F.S.L., Huffnagle W.O., Kowalik D.J., Lagrou M.,
Garber R.L., Goltzy L., Tolencino E., Westbrook-Wadman S., Yuan Y.,
Brodly L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
Kolzer J., Salter M.H., Hancock R.E.W., Lory S., Olson M.V.;

RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL: AE004962; AAG08886.1; -
DR InterPro: IPR001626; ABC-3.
DR Pfam: PF00950; ABC-3; 1.
KW Complete proteome.
SQ SEQUENCE 262 AA; 27343 MW; 95BF057DF127BA3E CRC64;

Query Match
Best Local Similarity 83.3%; Score 30; DB 2; Length 262;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxaxyxg 8
|| || |
Db 28 WRRMAYFG 35

RESULT 8
ID 064761 PRELIMINARY; PRT; 303 AA.
AC 064761;
DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
DE PUTATIVE PHOSPHATIDYLINOSITOL-GLYCAN-CLASS C (PGC).
GN F1913.21.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eupsidis II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,
Sykes S.M., Kaul S., Mason T.M., Kellavage A.R., Adams M.D.,
RA Somerville C.R., Venter J.C.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC004238; AAC12837.1; -
SQ SEQUENCE 303 AA; 34200 MW; 88E6BF82F4940C92 CRC64;

Query Match
Best Local Similarity 83.3%; Score 30; DB 10; Length 303;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxaxyxg 8
|| || |
Db 15 WRKVAYGG 22

RESULT 9
ID 09RD13 PRELIMINARY; PRT; 660 AA.
AC 09RD13;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE PUTATIVE REGULATOR.
GN SCML.31C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RD EMBL; AL133422; CAB62689.1; -
SQ SEQUENCE 660 AA; 70521 MW; EA9AC4C028FA8242 CRC64;

Query Match 83.3%; Score 30; DB 2; Length 660;
Best Local Similarity 62.5%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
DB 135 WRVHAYKG 142

RESULT 10
O9LDE7 PRELIMINARY; PRT; 938 AA.

DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE EST C28952(C62945) CORRESPONDS TO A REGION OF THE PREDICTED GENE.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nippobare(GA3) genomic DNA, chromosome 1, PAC
clone:p0699D11.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nippobare(GA3) genomic DNA, chromosome 1, PAC
clone:p0469E09.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RD EMBL; AP002817; BAB03449.1; -
RD EMBL; AP001366; BAA92402.1; -
DR InterPro; IPR001878; Znf_C2HC.
DR SMART; SMO0343; Znf_C2HC.1.
SQ SEQUENCE 938 AA; 108048 MW; B7689BD9E483C591 CRC64;

Query Match 83.3%; Score 30; DB 10; Length 938;
Best Local Similarity 62.5%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
DB 266 WRVHAYKG 273

RESULT 11
O9XE23 PRELIMINARY; PRT; 955 AA.
AC O9XE23;
DT 01-NOV-1999 (TREMBlrel. 12, Created)

DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE HYPOTHETICAL 109.6 KDA PROTEIN.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa Nippobare(GA3) genomic DNA, chromosome 2, PAC
clone:p0437H03 (contig b).";
RL Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
RD EMBL; AP000367; BAA82378.1; -
KW Hypothetical protein.
SQ SEQUENCE 955 AA; 109618 MW; B8DF7F0ED5112ECB CRC64;

Query Match 83.3%; Score 30; DB 10; Length 955;
Best Local Similarity 62.5%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
DB 115 WRVHAYKG 122

RESULT 12
O9FRA2 PRELIMINARY; PRT; 1011 AA.

DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE SIMILAR TO ORYZA SATIVA MUTATOR-LIKE TRANSPOSASE (AC068924).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Chow T.-Y., Hsing Y.-I.C., Chen H.-H., Wu H.-P., Chao Y.-T.,
RA Liu S.-M., Hsiao Y.-Y., Huang J.-J., Lee P.-F., Su C.-L., Chen C.-S.,
RA Shaw J.-F.;
RT "Oryza sativa PAC P0001A07 genomics sequence.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
RD EMBL; AC084218; AAG48840.1; -
SQ SEQUENCE 1011 AA; 113856 MW; EFE9AE49910C5281 CRC64;

Query Match 83.3%; Score 30; DB 10; Length 1011;
Best Local Similarity 62.5%; Pred. No. 2.1e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
DB 277 WRVHAYKG 284

RESULT 13
O9LDA3 PRELIMINARY; PRT; 1281 AA.
ID O9LDA3
AC O9LDA3;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE EST C28952(C62945) CORRESPONDS TO A REGION OF THE PREDICTED GENE.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Ehirharoldaeae; Oryzeae; Oryza.
OK NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:PO699D11.";
RN Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:PO469E09.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF002817; BAB03445.1; -;
DR EMBL: AF001366; BAA92398.1; -;
SQ SEQUENCE 1281 AA; 148721 MW; D6BE72AD3FA3DCC0 CRC64;

Query Match 83.3%; Score 30; DB 10; Length 1281;
Best Local Similarity 62.5%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
DB 237 WRVHAYKG 244

RESULT 14
OYLDW9 PRELIMINARY; PRT; 1591 AA.

AC Q9LDM9;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE EST C28952(C62945) CORRESPONDS TO A REGION OF THE PREDICTED GENE.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehirharoldaeae; Oryzeae; Oryza.
OK NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:PO469E05.";
RN Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:PO511C01.";
RN Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF002480; BAA96559.1; -;
DR EMBL: AF002070; BAA95853.1; -;
DR InterPro: IPR001878; Znf_CCHC.
DR SMART: SM00343; Znf_C2HC.1.
SQ SEQUENCE 1591 AA; 180064 MW; 461067C0252D4510 CRC64;

Query Match 83.3%; Score 30; DB 10; Length 1591;
Best Local Similarity 62.5%; Pred. No. 3.4e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
DB 266 WRVHAYKG 273

RESULT 15
O9FW81 PRELIMINARY; PRT; 1626 AA.

AC O9FW81;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE MUTATOR-LIKE TRANSPOSASE.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehirharoldaeae; Oryzeae; Oryza.
OK NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, NIPPONBARE;
RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Jenkins C.N., Burr P.C.,
RA Hsiao J., Zismann V., Pal G., Bowman C.L., Fujii C.Y., Vankken S.E.,
RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblum T.V.,
RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
RT "Oryza sativa chromosome 10 BAC OSJNBa0026L12 genomic sequence.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC068924; AAG13514.1; -;
DR InterPro: IPR001878; Znf_CCHC.
DR SMART: SM00343; Znf_C2HC.1.
SQ SEQUENCE 1626 AA; 183480 MW; A28A98C9BBA30075 CRC64;

Query Match 83.3%; Score 30; DB 10; Length 1626;
Best Local Similarity 62.5%; Pred. No. 3.4e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
DB 266 WRVHAYKG 273

RESULT 16
O9AYG3 PRELIMINARY; PRT; 1641 AA.

AC O9AYG3;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE MUTATOR-LIKE TRANSPOSASE.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehirharoldaeae; Oryzeae; Oryza.
OK NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA See L.H., Spiegel L.A., Nascimento L.U., de la Bastide M.,
RA Preston R.R., Huang E.N., Rodriguez M.A., Vil M.D., Baker J.P.,
RA Bahret A., Shah R.S., Miller B., Kirchoff K.A., King L., Toch K.,
RA O'Shaughnessy A., Dedhia N.N., McComble W.R.;
RT "Genomic Sequence for Oryza sativa, Nipponbare strain, Chromosome X,
Clone OSJNBa0094J09, complete sequence.";
RN Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RA McComble W.R.;
RN Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA McComble W.R.;
RN Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Bal H.P., See L.H., Spiegel L.A., Nascimento L.U., de la Bastide M.,
RA Preston R.R., Huang E.N., Rodriguez M.A., Vil M.D., Baker J.P.,
RA Bahret A., Shah R.S., Miller B., Kirchoff K.A., King L., Toch K.,
O'Shaughnessy A., Dedhia N.N., McComble W.R.;

RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC078839; AAK13091.1; -
SQ SEQUENCE 1641 AA; 184860 MW; 63C9BCB0EF1E6AC2 CRC64;

Query Match 83.3%; Score 30; DB 10; Length 1641;
Best Local Similarity 62.5%; Pred. No. 3.5e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
||:|
Db 266 WRVHAYKG 273

RESULT 17
O9RD02 PRELIMINARY; PRT; 65 AA.
AC O9RD02; 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE HYPOTHETICAL 6.9 KDA PROTEIN.
GN SCA47.10.

OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.

RC MEDLINE=97000351; Pubmed=8843436;
RX Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL133423; CAB62714.1; -.

KW Hypothetical protein.
SO SEQUENCE 65 AA; 6944 MW; F283FA15A0650DCE CRC64;

Query Match 80.6%; Score 29; DB 2; Length 65;
Best Local Similarity 50.0%; Pred. No. 21;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
||:|
Db 12 WRKSSYSG 19

RESULT 18
O9RD18 PRELIMINARY; PRT; 65 AA.
AC O9RD18; 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE HYPOTHETICAL 7.0 KDA PROTEIN.
GN SCC57A.09C.

OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE=97000351; Pubmed=8843436;
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL136519; CAB6277.1; -.
KW Hypothetical protein.
SO SEQUENCE 65 AA; 6959 MW; 1F74C265B9572610 CRC64;

Query Match 80.6%; Score 29; DB 2; Length 65;
Best Local Similarity 50.0%; Pred. No. 21;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
||:|
Db 7 WRSSYSG 14

RESULT 19
O9RD17 PRELIMINARY; PRT; 65 AA.
AC O9RD17; 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE HYPOTHETICAL 7.1 KDA PROTEIN.
GN SCC57A.10C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE=97000351; Pubmed=8843436;
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL136519; CAB6278.1; -.
KW Hypothetical protein.
SO SEQUENCE 65 AA; 7127 MW; 4C9A4476C44B727A CRC64;

Query Match 80.6%; Score 29; DB 2; Length 65;
Best Local Similarity 50.0%; Pred. No. 21;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxayxg 8
 ||:|
 Db 7 WRKSSYSG 14

RESULT 20

Q9KPP6 PRELIMINARY; PRT: 1208 AA.
 AC Q9KPP6;
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE EXODEOXYRIBONUCLEASE V, 135 KDA SUBUNIT.
 GN VC2320.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
 OX NCBI_Taxid 666;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN:EL TOR N16961 / SEROTYPE O1;
 MEDLINE:20406833; PubMed:10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
 Dodson R.D., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
 McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
 Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 cholerae."
 RL Nature 406:477-483(2000).
 DR EMBL: AE004302; AAF95464.1; .
 DR TIGR: VC2320; .
 DR InterPro: IPR000212; UvrD-helicase.
 DR Pfam: PF00580; UvrD-helicase; 2.
 KW Complete proteome.
 SO SEQUENCE 1208 AA; 135860 MW; F3C3EF891435C18D CRC64;

Query Match 80.6%; Score 29; DB 2; Length 1208;
 Best Local Similarity 50.0%; Pred. No. 4.2e+02;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 OY 1 wrxayxg 8
 ||:|
 Db 906 WRVTSYSG 913

Search completed: January 14, 2002, 07:39:47
 Job time: 958 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:36:01 : Search time 103.51 Seconds
(without alignments)
5.725 Million cell updates/sec

Title: 09-185908-1f
Perfect score: 36
Sequence: 1 wrxxxyxg 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database :

```
1: A_Geneseq_1101.*
2: /SID52/gcgdata/geneseq/geneseqp/AA1980.DAT.*
3: /SID52/gcgdata/geneseq/geneseqp/AA1981.DAT.*
4: /SID52/gcgdata/geneseq/geneseqp/AA1982.DAT.*
5: /SID52/gcgdata/geneseq/geneseqp/AA1983.DAT.*
6: /SID52/gcgdata/geneseq/geneseqp/AA1984.DAT.*
7: /SID52/gcgdata/geneseq/geneseqp/AA1985.DAT.*
8: /SID52/gcgdata/geneseq/geneseqp/AA1986.DAT.*
9: /SID52/gcgdata/geneseq/geneseqp/AA1987.DAT.*
10: /SID52/gcgdata/geneseq/geneseqp/AA1988.DAT.*
11: /SID52/gcgdata/geneseq/geneseqp/AA1989.DAT.*
12: /SID52/gcgdata/geneseq/geneseqp/AA1990.DAT.*
13: /SID52/gcgdata/geneseq/geneseqp/AA1991.DAT.*
14: /SID52/gcgdata/geneseq/geneseqp/AA1992.DAT.*
15: /SID52/gcgdata/geneseq/geneseqp/AA1993.DAT.*
16: /SID52/gcgdata/geneseq/geneseqp/AA1994.DAT.*
17: /SID52/gcgdata/geneseq/geneseqp/AA1995.DAT.*
18: /SID52/gcgdata/geneseq/geneseqp/AA1996.DAT.*
19: /SID52/gcgdata/geneseq/geneseqp/AA1997.DAT.*
20: /SID52/gcgdata/geneseq/geneseqp/AA1998.DAT.*
21: /SID52/gcgdata/geneseq/geneseqp/AA1999.DAT.*
22: /SID52/gcgdata/geneseq/geneseqp/AA2000.DAT.*
23: /SID52/gcgdata/geneseq/geneseqp/AA2001.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	88.9	496	22	AAG81335 Human AFP protein
2	32	88.9	496	22	AAB88466 Human membrane or
3	32	88.9	497	21	AA812138 Hydrophobic domain
4	31	86.1	513	11	AA804585 Aquaticus I. The
5	31	86.1	513	12	AA813181 T. aquaticus Aquari
6	31	86.1	513	16	AAR67653 Aquaflysin I. Ther
7	31	86.1	934	19	AAW82254 UP10246575 Seq ID
8	29	80.6	8	21	AAB06521 Claudin-2 cell adh
9	29	80.6	8	21	AAB06574 Claudin-2 cyclic c
10	29	80.6	10	21	AAB06530 Claudin-2 cyclic c
11	29	80.6	10	21	AAB06539 Claudin-2 cyclic c

12	29	80.6	10	21	AAB06548
13	29	80.6	10	21	AAB06557
14	29	80.6	10	21	AAB06566
15	29	80.6	230	10	AAV51676
16	29	80.6	448	22	AAB52467
17	28	77.8	8	21	AAB06426
18	28	77.8	8	21	AAB06512
19	28	77.8	9	21	AAB06427
20	28	77.8	10	21	AAB06485
21	28	77.8	10	21	AAB06491
22	28	77.8	10	21	AAB06497
23	28	77.8	10	21	AAB06503
24	28	77.8	10	21	AAB06509
25	28	77.8	71	20	AAV38421
26	28	77.8	211	20	AAV41726
27	28	77.8	211	20	AAV04143
28	28	77.8	211	20	AAW93653
29	28	77.8	211	21	AA844282
30	28	77.8	211	21	AAV66679
31	28	77.8	212	20	AAV38430
32	28	77.8	212	21	AAV76130
33	28	77.8	275	22	AA690433
34	27	75.0	8	21	AAB06583
35	27	75.0	8	21	AAB06636
36	27	75.0	8	21	AAB06645
37	27	75.0	8	21	AAB06658
38	27	75.0	10	21	AAB06592
39	27	75.0	10	21	AAB06601
40	27	75.0	10	21	AAB06610
41	27	75.0	10	21	AAB06619
42	27	75.0	10	21	AAB06628
43	27	75.0	10	21	AAB06654
44	27	75.0	10	21	AAB06663
45	27	75.0	10	21	AAB06672
46	27	75.0	10	21	AAB06681
47	27	75.0	10	21	AAB06690
48	27	75.0	202	20	AAW88629
49	27	75.0	205	21	AAW80533
50	27	75.0	208	21	AAB52100

ALIGNMENTS

RESULT	1
AA681335	
ID	AA681335 standard; Protein; 496 AA.
XX	
AC	AA681335;
XX	
DT	10-SEP-2001 (first entry)
XX	
DE	Human AFP protein sequence SEQ ID NO:188.
XX	
KW	Human; secreted protein; secretion; bacterial cell; fungal cell;
KW	eukaryotic cell; fusion protein; maltose binding protein;
KW	immunoglobulin constant region; polystyridine tag.
XX	
OS	Homo sapiens.
XX	
PN	W0200129221-A2.
XX	
PD	26-APR-2001.
XX	
PF	20-OCT-2000; 2000WO-US29052.
XX	
PR	20-OCT-1999; 99US-0160712.
XX	
PA	(ZYMO) ZYMOGENETICS INC.
XX	
PI	Conklin DC, Yee DP;
XX	
DR	WPI; 2001-300340/31.

Claudin-2 cyclic c
Claudin-2 cyclic c
Claudin-2 cyclic c
Murine clodin 2 pr
Mycobacterium tube
Claudin-1 cell adh
Claudin-1 cyclic c
Claudin-1 cell adh
Claudin-1 cyclic c
Claudin-1 cyclic c
Human secreted pro
Human PRO944 prote
Human Tango-73 pro
Human senescence f
Human PRO944 (UNQ4
A human molecule a
Human secreted pro
Human secreted pro
C glutamicum prote
Claudin-3 cell adh
Claudin-3 cyclic c
Claudin-3 cyclic c
Claudin-3 cyclic c
Claudin-4 cyclic c
Claudin-4 cyclic c
Claudin-4 cyclic c
Claudin-4 cyclic c
Claudin-4 cyclic c
Claudin-4 cyclic c
Secreted protein e
Human ORF297
Gene 48 human secr

DR N-PSDB; AAH52186.
XX Isolated polypeptide for directing secretion of proteins of interest
XX from a host cell including, e.g. bacteria, includes contiguous amino
PT acid residues of polypeptide with specified amino acids
XX
XX Claim 1; Page 337-338; 617pp; English.
XX
CC AAH52093 to AAH52303 encode the human secreted proteins given in AAG81242
CC to AAG81453. The secreted proteins can be used for directing the
CC secretion of proteins of interest from a host cell including bacteria,
CC fungal cells, and cultured higher eukaryotic cells. The present invention
CC also describes fusion proteins, where a secreted protein of the invention
CC is operably linked via a peptide bond or peptide linker to a second
CC protein selected from the group consisting of maltose binding protein,
CC an immunoglobulin constant region, a polyhistidine tag and a peptide
CC given in AAG81453.
XX
SQ Sequence 496 AA;

Query Match 88.9%; Score 32; DB 22; Length 496;
Best Local Similarity 62.5%; Pred. No. 55;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxayxg 8
|| || |
Db 4 wrgsayag 11

RESULT 2
AAB88466
ID AAB88466 standard; Protein: 496 AA.
XX
AC AAB88466;
XX
DT 23-MAY-2001 (first entry)
XX
DE Human membrane or secretory protein clone PSEC0260.
XX
KW Human; secretory protein; membrane protein; vaccine; gene therapy;
KW rheumatoid arthritis; diabetes.
XX
OS Homo sapiens.
XX
PN EPI067182-A2.
XX
PD 10-JAN-2001.
XX
PF 07-JUL-2000; 2000EP-0114090.
XX
PR 08-JUL-1999; 99JP-0194179.
PR 11-JAN-2000; 2000JP-0118775.
PR 02-MAY-2000; 2000JP-0183766.
XX
PA (HELI-) HELIX RES INST.
PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX WPI: 2001-093989/11.
DR N-PSDB; AAF93893.
XX
XX Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT gene therapy or as candidate target molecules in drug development -
XX
PS Claim 1; SEQ ID 300; 609pp + CD ROM; English.
XX
CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC which encode human secretory or membrane proteins represented by
CC AA984317 - AAB88419. Included in the invention are primers
CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC cDNA sequences of the invention. The invention also includes methods for
CC the production of antibodies directed against the proteins, and cDNA

CC sequences, which can be used in vaccines. The polynucleotide sequences
CC can be used in gene therapy. The polynucleotide sequences and the
CC proteins they encode may be used in the prevention, treatment and
CC diagnosis of diseases associated with inappropriate secretory
CC protein/membrane protein expression. The nucleic acids and complementary
CC sequences may also be used as DNA probes in diagnostic assays
CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC presence of similar nucleic acid sequences in samples. They may also be
CC used to study the expression and function of secretory proteins/membrane
CC polypeptides and their role in metabolism. The polypeptides may be used
CC as antigens in the production of antibodies against them and in assays to
CC identify modulators (agonists and antagonists) of expression and
CC activity. The antibodies and antagonists may also be used as therapeutic
CC agents to down regulate expression and activity. The antibodies may also
CC be used as diagnostic agents for detecting the presence of the
CC polypeptides in samples (e.g. by enzyme linked immunosorbant assay
CC (ELISA). Examples of diseases which may be treated include rheumatoid
CC arthritis and diabetes.
XX
SQ Sequence 496 AA;

Query Match 88.9%; Score 32; DB 22; Length 496;
Best Local Similarity 62.5%; Pred. No. 55;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxayxg 8
|| || |
Db 4 wrgsayag 11

RESULT 3
AAB12138
ID AAB12138 standard; Protein: 497 AA.
XX
AC AAB12138;
XX
DT 02-FEB-2001 (first entry)
XX
DE Hydrophobic domain protein isolated from HT-1080 cells.
XX
KW Human; secreted protein; membrane protein; hydrophobic domain;
KW proliferation control; differentiation induction; material transport;
KW biophylaxis; signal receptor; ion channel; transporter; immunostimulant;
KW immunosuppressant; haematopoiesis regulator; chemotactic; chemokinetic;
KW haemostatic; thrombolytic; anti-inflammatory; tumour inhibition;
KW autoimmune disease; Alzheimer's disease; Parkinson's disease; cancer.
XX
OS Homo sapiens.
XX
PN WO200029448-A2.
XX
PD 25-MAY-2000.
XX
PF 17-NOV-1999; 99WO-JP06412.
XX
PR 17-NOV-1998; 98JP-0326255.
PR 22-DEC-1998; 98JP-0364315.
PR 16-MAR-1999; 99JP-0069811.
PR 27-APR-1999; 99JP-0119299.
PR 19-MAY-1999; 99JP-0138169.
XX
PA (SAGA) SAGAMI CHEM RES CENT.
PA (PROT-) PROTEGENE INC.
PI Kato S, Kimura T;
XX
XX WPI: 2000-387753/33.
DR N-PSDB; AAA62021, AAA62031.
XX
PT Proteins comprising hydrophobic regions, such as secretory and membrane
PT proteins, useful in research and diagnostics and having various
PT activities e.g. immunomodulatory, antiinflammatory, chemokinetic,

PR hemostatic, thrombolytic -
XX
PS Claim 1; Page 286-288; 410pp; English.
XX
CC Secretory proteins play important roles in the proliferation control, the
CC differentiation induction, the material transport and the biophysics of
CC cells. Membrane proteins have important roles as signal receptors, ion
CC channels and transporters. The present sequence is a human protein which
CC has at least one hydrophobic domain. This protein may be a secretory or a
CC membrane protein. The present protein may have cytokine and cell
CC proliferation/differentiation activity, immune stimulating or suppressing
CC activity, haematopoiesis activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, anti-inflammatory activity and tumour
CC inhibition activity. The present protein could therefore be used for
CC treatment of autoimmune disease, Alzheimer's disease, Parkinson's
CC disease, and cancer.
XX
SQ Sequence 497 AA:

Query Match 88.9%; Score 32; DB 21; Length 497;
Best Local Similarity 62.5%; Pred. No. 55;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxayxg 8
II III
DB 4 wrgsayyg 11

RESULT 4
AAK04585
ID AAK04585 standard; protein: 513 AA.
AC AAR04585;
XX
DT 24-SEP-1989 (first entry)
XX
DE Aquaricine I.
XX
KW Aquaricine I; Colibacillus; detergents; proteolysis.
XX
OS Thermus aquaticus.
XX
FH Key Location/Qualifiers
FT Peptide 1..22
FT /label=signal peptide
FT 23..127
FT Region /label=pro-Aquaricine I region
FT 128..408
FT /label=mature Aquaricine I
FT 409..513
FT /label=tail region
XX
PN JP02092288-A.
XX
PD 03-APR-1990.
XX
PF 30-JUL-1988; 88JP-0243981.
XX
PR 30-JUL-1988; 88JP-0243981.
XX
PA (NISB) JAPAN TOBACCO AND SALT PUB.
XX
DR MPI: 1990-144901/19.
DR N-PSDB; AAQ04339.
XX
PT Gene to code precursor of Aquaricine I-and Colibacillus contg. expression
PT vector, for proteolytic enzyme produ. for detergent additive
PS Disclosure; ; p: Japanese.
XX
CC The Aquaricine I precursor is expressed in a Colibacillus host. The

CC mature protein is recovered following processing.
CC Aquaricine I is a heat resistant proteolytic enzyme, used eg as an addit-
CC ive in detergents and as a catalyst for peptide synthesis.
XX
SQ Sequence 513 AA:

Query Match 86.1%; Score 31; DB 11; Length 513;
Best Local Similarity 62.5%; Pred. No. 92;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxayxg 8
II III
DB 495 wriyaysg 502

RESULT 5
AAR13181
ID AAR13181 standard; protein: 513 AA.
XX
AC AAR13181;
XX
DT 11-OCT-1991 (first entry)
XX
DE T.aquaticus Aquaricin 1.
XX
KW thermophilic bacteria; Thermus thermophilus HB27; ss.
XX
OS Thermus aquaticus YT-1.
XX
FH Key Location/Qualifiers
FT Peptide 1..22
FT /label=signal_peptide
FT 23..127
FT Region /label=pro_region
FT 128..408
FT Protein /label=Aquaricin 1
FT 409..513
FT Region /label=tail region
XX
PN JP03151880-A.
XX
PD 28-JUN-1991.
XX
PF 08-NOV-1989; 89JP-0288813.
XX
PR 08-NOV-1989; 89JP-0288813.
XX
PA (NISB) JAPAN TOBACCO INC.
XX
DR MPI: 1991-234066/32.
DR N-PSDB; AAQ12838.
XX
PT Expression vector contains aquaricin I coding gene - in extremely
PT thermophilic gram negative bacteria
XX
PS Disclosure; Fig 1; 10pp; Japanese.
XX
CC The sequence encoding this protein was derived from T.aquaticus
CC YT-1. Aquaricin 1 is secreted from T.thermophilus HB27 transformed
CC with a vector (pMK006) containing the coding sequence.
XX
SQ Sequence 513 AA:

Query Match 86.1%; Score 31; DB 12; Length 513;
Best Local Similarity 62.5%; Pred. No. 92;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxayxg 8
II III
DB 495 wriyaysg 502

RESULT 6
AAR67653
ID AAR67653 standard; Protein: 513 AA.
AC AAR67653;
XX
DT 16-AUG-1995 (first entry)
XX
DE Aqualysin I.
XX
KW Aqualysin I; glucomylase; extracellular production.
XX
OS Thermus aquaticus.
XX
FH Key Location/Qualifiers
FT Peptide 1..127 /label= sig_peptide
XX
PN .JP06303984-A.
XX
PD 01-NOV-1994.
XX
PF 19-APR-1993; 93JP-0115378.
XX
PR 19-APR-1993; 93JP-0115378.
XX
PA (NISB) JAPAN TOBACCO INC.
XX
DR WPI: 1995-018277/03.
XX N-PSDB: AAO75859.
XX
PT Recombinant DNA having sequence deleted from the aqua-lysin I
XX precursor gene - for expression of mature aqua-lysin I in yeast
PS Claim 1: Pages 8-10; 16pp; Japanese.
XX
CC AAO75858 is the yeast glucomylase (AAR67652) DNA, into which the
CC aqualysin I (AAR67653) DNA (AAO75859) minus its C-terminal pro
CC sequence can be inserted. Resulting in the extracellular
CC production of aqualysin I in yeast.
XX
SQ Sequence 513 AA;

Query Match 86.1%; Score 31; DB 16; Length 513;
Best Local Similarity 62.5%; Pred. No. 92;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxayxg 8
|| || |
DB 495 wrlyayysg 502

RESULT 7
AAW82254
ID AAW82254 standard; Protein: 934 AA.
XX
AC AAW82254;
XX
DT 16-JUL-1999 (first entry)
XX
DE JP10248575 Seq ID 4.
XX
KW HMG-CoA: 3-hydroxy-3-methylglutaryl coenzyme A reductase; carotenoid.
XX
OS Unidentified.
XX
PN JP10248575-A.
XX
PD 22-SEP-1998.
XX
PF 12-MAR-1997; 97JP-0058012.

XX
PR 12-MAR-1997; 97JP-0058012.
XX
PA (KIRI) KIRIN BREWERY KK.
XX
DR WPI: 1998-560727/48.
XX N-PSDB: AAV73463.
XX
PT Gene useful for increase in carotenoid production - and preparation
XX of carotenoid
XX
PS Disclosure; Page 18-19; 54pp; Japanese.
XX
CC This invention describes a novel method for the preparation of
CC carotenoids using genes and proteins isolated from Candida utilis.
CC This sequence represents a protein which is used in the method of the
XX invention.
XX
SQ Sequence 934 AA;

Query Match 86.1%; Score 31; DB 19; Length 934;
Best Local Similarity 62.5%; Pred. No. 1,6e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxayxg 8
|| || |
DB 79 wrstrayng 86

RESULT 8
AAB06521
ID AAB06521 standard; peptide: 8 AA.
XX
AC AAB06521;
XX
DT 28-SEP-2000 (first entry)
XX
DE Claudin-2 cell adhesion recognition sequence SEQ ID NO: 42.
XX
KW Claudin-2 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection.
XX
OS Mammalia.
XX
PN WO200026360-A1.
XX
PD 11-MAY-2000.
XX
PF 03-NOV-1999; 99WO-CA01029.
XX
PR 03-NOV-1998; 98US-0185908.
XX PR 30-MAR-1999; 99US-0282029.
XX
PA (ADHE-) ADHEREX TECHNOLOGIES INC.
XX
PI Blaschuck OW, Symonds JM, Gour BJ;
XX
DR WPI: 2000-365610/31.
XX
PT Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
PS Claim 46; Page 97; 121pp; English.
XX
CC The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-2 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant

CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site.

XX
SQ Sequence 8 AA:

Query Match 80.6%; Score 29; DB 21; Length 8;
Best Local Similarity 50.0%; Pred. No. 4.3e+05;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
11:111
DB 1 wrtssyvg 8

RESULT 9

AAB06574
ID AAB06574 standard; peptide: 8 AA.

XX AAB06574;

XX 28-SEP-2000 (first entry)

XX claudin-2 cyclic cell adhesion recognition sequence SEQ ID NO: 163.

XX claudin-2 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection; cyclic.

XX Mammalia.

XX WO200026360-A1.

XX 11-MAY-2000.

XX 03-NOV-1999; 99WO-CA01029.

XX 03-NOV-1998; 98US-0185908.

XX 30-MAR-1999; 99US-0282029.

XX (ADHE-) ADHEREX TECHNOLOGIES INC.

XX Blaschuck OW, Symonds JM, Gour BJ;

XX WPI: 2000-365610/31.

XX Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -

XX Claim 49; Page 98; 121pp; English.

XX The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-2 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.

XX Sequence 8 AA:

Query Match 80.6%; Score 29; DB 21; Length 8;
Best Local Similarity 50.0%; Pred. No. 4.3e+05;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
11:111
DB 1 wrtssyvg 8

RESULT 10

AAB06530
ID AAB06530 standard; peptide: 10 AA.

XX AAB06530;

XX 28-SEP-2000 (first entry)

XX claudin-2 cyclic cell adhesion recognition sequence SEQ ID NO: 119.

XX claudin-2 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection; cyclic.

XX Mammalia.

XX WO200026360-A1.

XX 11-MAY-2000.

XX 03-NOV-1999; 99WO-CA01029.

XX 03-NOV-1998; 98US-0185908.

XX 30-MAR-1999; 99US-0282029.

XX (ADHE-) ADHEREX TECHNOLOGIES INC.

XX Blaschuck OW, Symonds JM, Gour BJ;

XX WPI: 2000-365610/31.

XX Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -

XX Claim 49; Page 98; 121pp; English.

XX The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-2 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.

XX Sequence 10 AA:

Query Match 80.6%; Score 29; DB 21; Length 10;
Best Local Similarity 50.0%; Pred. No. 5.6;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
11:111
DB 2 wrtssyvg 9

RESULT 11

AAB06539
ID AAB06539 standard; peptide: 10 AA.

XX AAB06539;

XX 28-SEP-2000 (first entry)

XX claudin-2 cyclic cell adhesion recognition sequence SEQ ID NO: 128.

XX claudin-2 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection; cyclic.

```
XX OS Mammalia.
XX PN WO200026360-A1.
XX PD 11-MAY-2000.
XX PF 03-NOV-1999; 99WO-CA01029.
XX PR 03-NOV-1998; 98US-0185908.
XX PR 30-MAR-1999; 99US-0282029.
XX PA (ADHE-) ADHEREX TECHNOLOGIES INC.
XX PI Blaschuck OW, Symonds JM, Gour BJ;
XX DR WPI: 2000-365610/31.
XX PT Antibody modulation of claudin-mediated cell adhesion for increasing
XX PT vasopermeability, for delivering drugs to tumors and the nervous system
XX PT and across the skin -
XX PS Claim 49; Page 98; 121pp; English.
XX CC The present invention relates to the use of peptides as claudin-mediated
XX CC cell adhesion modulators. The claudin-2 group of proteins are cadherins,
XX CC which are membrane glycoproteins involved in cell adhesion. In some
XX CC situations, cell adhesion occurs at abnormal levels, and these peptides
XX CC can be used to modulate these levels, and thus treat autoimmune diseases,
XX CC inflammatory diseases and cancer, and aid wound healing and implant
XX CC adhesion. In addition, they can also be used to facilitate drug delivery
XX CC to the desired target site. The present sequence has a cyclic
XX CC conformation.
XX SQ Sequence 10 AA;

Query Match 80.6%; Score 29; DB 21; Length 10;
Best Local Similarity 50.0%; Pred. No. 5.6;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 wrxxayxg 8
Db ||:| |
2 wrtssyvg 9

RESULT 12
AAB06548
ID AAB06548 standard; peptide: 10 AA.
XX AC AAB06548;
XX DT 28-SEP-2000 (first entry)
XX DE Claudin-2 cyclic cell adhesion recognition sequence SEQ ID NO: 137.
XX KW Claudin-2 modulating agent; cell adhesion recognition sequence;
XX KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
XX KW graft rejection; cyclic.
XX OS Mammalia.
XX PN WO200026360-A1.
XX PD 11-MAY-2000.
XX PF 03-NOV-1999; 99WO-CA01029.
XX PR 03-NOV-1998; 98US-0185908.
XX PR 30-MAR-1999; 99US-0282029.
XX PA (ADHE-) ADHEREX TECHNOLOGIES INC.
XX PI Blaschuck OW, Symonds JM, Gour BJ;
XX DR WPI: 2000-365610/31.
XX PT Antibody modulation of claudin-mediated cell adhesion for increasing
XX PT vasopermeability, for delivering drugs to tumors and the nervous system
XX PT and across the skin -
XX PS Claim 49; Page 98; 121pp; English.
XX CC The present invention relates to the use of peptides as claudin-mediated
XX CC cell adhesion modulators. The claudin-2 group of proteins are cadherins,
XX CC which are membrane glycoproteins involved in cell adhesion. In some
XX CC situations, cell adhesion occurs at abnormal levels, and these peptides
XX CC can be used to modulate these levels, and thus treat autoimmune diseases,
XX CC inflammatory diseases and cancer, and aid wound healing and implant
XX CC adhesion. In addition, they can also be used to facilitate drug delivery
XX CC to the desired target site. The present sequence has a cyclic
XX CC conformation.
XX SQ Sequence 10 AA;
```

```
PI Blaschuck OW, Symonds JM, Gour BJ;
XX DR WPI: 2000-365610/31.
XX PT Antibody modulation of claudin-mediated cell adhesion for increasing
XX PT vasopermeability, for delivering drugs to tumors and the nervous system
XX PT and across the skin -
XX PS Claim 49; Page 98; 121pp; English.
XX CC The present invention relates to the use of peptides as claudin-mediated
XX CC cell adhesion modulators. The claudin-2 group of proteins are cadherins,
XX CC which are membrane glycoproteins involved in cell adhesion. In some
XX CC situations, cell adhesion occurs at abnormal levels, and these peptides
XX CC can be used to modulate these levels, and thus treat autoimmune diseases,
XX CC inflammatory diseases and cancer, and aid wound healing and implant
XX CC adhesion. In addition, they can also be used to facilitate drug delivery
XX CC to the desired target site. The present sequence has a cyclic
XX CC conformation.
XX SQ Sequence 10 AA;

Query Match 80.6%; Score 29; DB 21; Length 10;
Best Local Similarity 50.0%; Pred. No. 5.6;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 wrxxayxg 8
Db ||:| |
2 wrtssyvg 9

RESULT 13
AAB06557
ID AAB06557 standard; peptide: 10 AA.
XX AC AAB06557;
XX DT 28-SEP-2000 (first entry)
XX DE Claudin-2 cyclic cell adhesion recognition sequence SEQ ID NO: 146.
XX KW Claudin-2 modulating agent; cell adhesion recognition sequence;
XX KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
XX KW graft rejection; cyclic.
XX OS Mammalia.
XX PN WO200026360-A1.
XX PD 11-MAY-2000.
XX PF 03-NOV-1999; 99WO-CA01029.
XX PR 03-NOV-1998; 98US-0185908.
XX PR 30-MAR-1999; 99US-0282029.
XX PA (ADHE-) ADHEREX TECHNOLOGIES INC.
XX PI Blaschuck OW, Symonds JM, Gour BJ;
XX DR WPI: 2000-365610/31.
XX PT Antibody modulation of claudin-mediated cell adhesion for increasing
XX PT vasopermeability, for delivering drugs to tumors and the nervous system
XX PT and across the skin -
XX PS Claim 49; Page 98; 121pp; English.
XX CC The present invention relates to the use of peptides as claudin-mediated
XX CC cell adhesion modulators. The claudin-2 group of proteins are cadherins,
XX CC which are membrane glycoproteins involved in cell adhesion. In some
XX CC situations, cell adhesion occurs at abnormal levels, and these peptides
```

CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.

XX Sequence 10 AA:

Query Match 80.6%; Score 29; DB 21; Length 10;
Best Local Similarity 50.0%; Pred. No. 5.6;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Oy 1 wrxxayxg 8
11:| |
Db 2 wrtsayvg 9

RESULT 14

AAB06566 AAB06566 standard; peptide: 10 AA.

AC AAB06566;

DT 28-SEP-2000 (first entry)

DE Claudin-2 cyclic cell adhesion recognition sequence SEQ ID NO: 155.

XX Claudin-2 mediating agent; cell adhesion recognition sequence;

KW CAR sequence; autoimmune disease; inflammatory disease; cancer;

KW graft rejection; cyclic.

XX Mammalia.

XX WO200026360-A1.

PD 11-MAY-2000.

PF 03-NOV-1999; 99MO-CA01029.

XX 03-NOV-1998; 98US-0185908.

PR 30-MAR-1999; 99US-0282029.

XX (ADHE-) ADHEREX TECHNOLOGIES INC.

PI Blaschuck OW, Symonds JM, Gour BJ;

XX WPI: 2000-365610/31.

PS Claim 49; Page 98; 121pp; English.

XX The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-2 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.

XX Sequence 10 AA:

Query Match 80.6%; Score 29; DB 21; Length 10;
Best Local Similarity 50.0%; Pred. No. 5.6;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 wrxxayxg 8

Db 11:| |
2 wrtsayvg 9

RESULT 15

AAV51676 AAV51676 standard; Protein: 230 AA.

XX AAV51676;

AC AAV51676;

DT 02-JUN-2000 (first entry)

DE Murine clodin 2 protein.

XX Clodin 2; murine; tight junction-constituting membrane protein;

KW medicine.

XX Mus sp.

XX JP2000032984-A.

PD 02-FEB-2000.

PF 26-JUN-1998; 98JP-0179847.

XX 15-MAY-1998; 98JP-0133215.

PA (EISA) EISAI CO LTD.

DR WPI: 2000-285512/25.

DR N-PSDB: AA289137.

XX Tight junction-constituting membrane protein clodin family - useful in

PT the medical field

XX Claim 2; Page 10; 22pp; Japanese.

XX This invention describes novel murine nucleic acid sequences encoding the

CC clodin family of tight junction (TJ)-constituting membrane protein. The

CC membrane protein can be used in medical field. This sequence represents

XX the clodin 2 protein described in the method of the invention.

XX Sequence 230 AA:

Oy 1 wrxxayxg 8
11:| |
Db 30 wrtsayvg 37

RESULT 16

AAB52467 AAB52467 standard; protein: 448 AA.

XX AAB52467;

DT 23-FEB-2001 (first entry)

DE Mycobacterium tuberculosis secreted protein #32.

XX Mycobacterium tuberculosis secreted protein; MTP; vaccine.

XX Mycobacterium tuberculosis.

XX WO20006143-A1.

PN 09-NOV-2000.

PF 04-MAY-2000; 2000WO-US12197.

XX 04-MAY-1999; 99US-0132479.
PR 04-MAY-1999; 99US-0132503.
XX
PA (PUBL-) PUBLIC HEALTH RES INST NEW YORK.
PI Gennaro ML, Gomez MJ;
XX WPI; 2001-007151/01.
XX
PT Novel Mycobacterium tuberculosis secreted polypeptides and
PT polynucleotides useful in diagnosis, treatment and prophylaxis of
XX tuberculosis -
XX
PS Claim 11; Fig 1; 60pp; English.
XX
CC The present invention relates to Mycobacterium tuberculosis secreted
CC proteins (MSP), where the polypeptide has M. tuberculosis specific
CC antigenic and immunogenic properties. Compositions of the invention may
CC be useful for diagnosing Mycobacterium tuberculosis infection and as a
CC vaccine against M. tuberculosis infection.
XX
SQ Sequence 448 AA;

Query Match 80.6%; Score 29; DB 22; Length 448;
Best Local Similarity 50.0%; Pred. No. 2.1e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 wrxxayxg 8
||:| |
Db 306 wrasgyysg 313

RESULT 17
AAB06426
ID AAB06426 standard; peptide; 8 AA.
XX
AC AAB06426;
XX
DT 28-SEP-2000 (first entry)
XX
DE Claudin-1 cell adhesion recognition sequence SEQ ID NO: 485.
XX
KW Claudin-1 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection.
XX
OS Mammalia.
XX
PN WO200026360-A1.
XX
PD 11-MAY-2000.
XX
PE 03-NOV-1999; 99WO-CA01029.
XX
PR 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
XX
PA (ADHE-) ADHEREX TECHNOLOGIES INC.
XX
PI Blaschuck OW, Symonds JM, Gour BJ;
XX WPI; 2000-365610/31.
XX
PT Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
PS Claim 39; Page 96; 121pp; English.
XX
CC The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-1 group of proteins are cadherins,
CC cell adhesion modulators. The claudin-1 group of proteins are cadherins,

CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site.
XX
SQ Sequence 8 AA;

Query Match 77.8%; Score 28; DB 21; Length 8;
Best Local Similarity 50.0%; Pred. No. 4.3e+05;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxayxg 8
||:| |
Db 1 wrlysyag 8

RESULT 18
AAB06512
ID AAB06512 standard; peptide; 8 AA.
XX
AC AAB06512;
XX
DT 28-SEP-2000 (first entry)
XX
DE Claudin-1 cyclic cell adhesion recognition sequence SEQ ID NO: 519.
XX
KW Claudin-1 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection; cyclic.
XX
OS Mammalia.
XX
PN WO200026360-A1.
XX
PD 11-MAY-2000.
XX
PE 03-NOV-1999; 99WO-CA01029.
XX
PR 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
XX
PA (ADHE-) ADHEREX TECHNOLOGIES INC.
XX
PI Blaschuck OW, Symonds JM, Gour BJ;
XX WPI; 2000-365610/31.
XX
PT Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
PS Claim 43; Page 97; 121pp; English.
XX
CC The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-1 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.
XX
SQ Sequence 8 AA;

Query Match 77.8%; Score 28; DB 21; Length 8;
Best Local Similarity 50.0%; Pred. No. 4.3e+05;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
11 : 11
DB 1 wrlysyag 8

RESULT 19
AAB06427
ID AAB06427 standard; peptide: 9 AA.
XX
AC AAB06427;
XX
DT 28-SEP-2000 (first entry)
XX
DE Claudin-1 cell adhesion recognition sequence SEQ ID NO: 486.
XX
KM Claudin-1 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
graft rejection.
OS Mammalia.
XX
PN WO200026360-A1.
XX
PD 11-MAY-2000.
XX
PF 03-NOV-1999; 99WO-CA01029.
XX
PR 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
XX
PA (ADHE-) ADHEREX TECHNOLOGIES INC.
XX
PI Blaschuck OW, Symonds JM, Gour BJ;
XX
DR WPI: 2000-365610/31.
XX
PT Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
PS Claim 39; Page 96; 121pp; English.
XX
CC The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-1 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site.
XX
SQ Sequence 9 AA:

Query Match 77.8%; Score 28; DB 21; Length 9;
Best Local Similarity 50.0%; Pred. NO. 4.3e+05;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
11 : 11
DB 1 wrlysyag 8

RESULT 20
AAB06485
ID AAB06485 standard; peptide: 10 AA.
XX
AC AAB06485;
XX
DT 28-SEP-2000 (first entry)
XX
DE Claudin-1 cyclic cell adhesion recognition sequence SEQ ID NO: 492.
XX

KW Claudin-1 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection; cyclic.
XX
OS Mammalia.
XX
PN WO200026360-A1.
XX
PD 11-MAY-2000.
XX
PF 03-NOV-1999; 99WO-CA01029.
XX
PR 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
XX
PA (ADHE-) ADHEREX TECHNOLOGIES INC.
XX
PI Blaschuck OW, Symonds JM, Gour BJ;
XX
DR WPI: 2000-365610/31.
XX
PT Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
PS Claim 43; Page 96; 121pp; English.
XX
CC The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-1 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.
XX
SQ Sequence 10 AA:

Query Match 77.8%; Score 28; DB 21; Length 10;
Best Local Similarity 50.0%; Pred. NO. 8.9;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxayxg 8
11 : 11
DB 2 wrlysyag 9

Search completed: January 14, 2002, 07:36:01
Job time: 812 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:23:44 ; Search time 48.05 Seconds
(without alignments)
3.747 Million cell updates/sec

Title: 09-185908-1f
Sequence: 1 wrxxxxyxg 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : Issued_Patents_AA: *
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep: *
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep: *
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep: *
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep: *
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep: *
6: /cgn2_6/prodata/2/1aa/backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	75.0	422	2 US-08-484-575A-6	Sequence 6, Appl1
2	27	75.0	422	3 US-08-477-159-6	Sequence 6, Appl1
3	27	75.0	422	3 US-08-479-869-6	Sequence 6, Appl1
4	27	75.0	422	4 US-08-486-414-6	Sequence 6, Appl1
5	27	75.0	422	5 PCT-US94-01826A-6	Sequence 6, Appl1
6	27	75.0	422	5 PCT-US94-02252A-6	Sequence 6, Appl1
7	27	75.0	1289	1 US-07-876-280-4	Sequence 4, Appl1
8	27	75.0	1289	1 US-07-875-772-4	Sequence 4, Appl1
9	27	75.0	1289	1 US-08-063-170-4	Sequence 4, Appl1
10	27	75.0	1289	1 US-08-158-332-4	Sequence 4, Appl1
11	27	75.0	1289	1 US-08-304-626-4	Sequence 4, Appl1
12	27	75.0	1289	1 US-08-316-801A-4	Sequence 4, Appl1
13	27	75.0	1289	2 US-08-611-928-4	Sequence 4, Appl1
14	27	75.0	1289	3 US-09-173-891-4	Sequence 4, Appl1
15	27	75.0	1289	4 US-09-076-137-4	Sequence 4, Appl1
16	27	75.0	1289	5 PCT-US92-03624-4	Sequence 4, Appl1
17	26	72.2	1220	1 US-08-158-232-43	Sequence 43, Appl1
18	26	72.2	1220	2 US-08-611-928-43	Sequence 43, Appl1
19	26	72.2	1220	3 US-09-173-891-43	Sequence 43, Appl1
20	26	72.2	1289	6 5281530-1	Patent No. 5281530
21	26	72.2	1289	6 5426049-4	Patent No. 5426049
22	26	72.2	1289	6 US-07-876-280-2	Sequence 2, Appl1
23	26	72.2	1289	1 US-07-675-772-2	Sequence 2, Appl1
24	26	72.2	1289	1 US-08-063-170-2	Sequence 2, Appl1
25	26	72.2	1289	1 US-08-158-232-2	Sequence 2, Appl1
26	26	72.2	1289	1 US-08-304-626-2	Sequence 2, Appl1
27	26	72.2	1289	1 US-08-316-801A-2	Sequence 2, Appl1

28	26	72.2	1385	2 US-08-611-928-2	Sequence 2, Appl1
29	26	72.2	1385	3 US-09-173-891-2	Sequence 2, Appl1
30	26	72.2	1385	4 US-09-076-137-2	Sequence 2, Appl1
31	26	72.2	1385	5 PCT-US92-03624-2	Sequence 2, Appl1
32	26	72.2	1385	6 5281530-1	Patent No. 5281530
33	26	72.2	1385	6 5426049-1	Patent No. 5426049
34	26	72.2	1620	1 US-08-542-363-2	Sequence 2, Appl1
35	26	72.2	1620	4 US-09-100-089-2	Sequence 2, Appl1
36	25	69.4	432	2 US-08-677-049-8	Sequence 8, Appl1
37	24	66.7	164	2 US-08-911-319A-1	Sequence 1, Appl1
38	24	66.7	164	3 US-09-352-619-1	Sequence 1, Appl1
39	24	66.7	376	1 US-08-608-241-2	Sequence 2, Appl1
40	24	66.7	376	2 US-08-922-182-2	Sequence 2, Appl1
41	24	66.7	376	2 US-08-918-953-2	Sequence 2, Appl1
42	24	66.7	376	4 US-09-192-983-2	Sequence 2, Appl1
43	24	66.7	748	1 US-08-408-318-2	Sequence 2, Appl1
44	24	66.7	748	1 US-08-369-796-10	Sequence 10, Appl1
45	24	66.7	748	1 US-08-839-164-2	Sequence 2, Appl1
46	24	66.7	748	2 US-08-852-091-10	Sequence 10, Appl1
47	24	66.7	748	2 US-08-820-734-10	Sequence 10, Appl1
48	24	66.7	748	3 US-08-956-652-10	Sequence 10, Appl1
49	24	66.7	748	3 US-08-956-869-10	Sequence 10, Appl1
50	24	66.7	748	3 US-08-948-547-10	Sequence 10, Appl1

ALIGNMENTS

RESULT 1
US-08-484-575A-6
Sequence 6, Application US/08484575A
Patent No. 5925358
GENERAL INFORMATION:
APPLICANT: Mark D. Cochran and David E. Junker
TITLE OF INVENTION: Recombinant Fowlpox Viruses and Uses Thereof
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESSES:
ADDRESSEE: John P. White
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484, 575A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 42A
ATTORNEY/AGENT INFORMATION:
NAME: White Esq, John P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)278-0450
TELEFAX: (212)391-0525
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 422 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-575A-6
Query Match
Best Local Similarity 75.0%; Score 27; DB 2; Length 422;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 wrxxxxyxg 8
Db 5 WKRTYSG 12

RESULT 2
US-08-477-459-6
Sequence 6, Application US/08477459
Patent No. 6001369
GENERAL INFORMATION:
APPLICANT: Mark D. Cochran
TITLE OF INVENTION: Recombinant Fowlpox Viruses and Uses
TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,459
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: White Esq, John P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 422 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-477-459-6

Query Match 75.0%; Score 27; DB 3; Length 422;
Best Local Similarity 50.0%; Pred. No. 2e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wxxxayxg 8
|| | |
Db 5 WRTEIYSG 12

RESULT 3
US-08-479-869-6
Sequence 6, Application US/08479869
Patent No. 6123949
GENERAL INFORMATION:
APPLICANT: Cochran Ph.D, Mark D
TITLE OF INVENTION: Recombinant Fowlpox Virus S-FPV-043 and
TITLE OF INVENTION: Uses Thereof
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,869

FILING DATE: 424
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/024,156
FILING DATE: 26-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: White Esq, John P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)977-9550
TELEFAX: (212)664-0525
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 422 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-479-869-6

Query Match 75.0%; Score 27; DB 3; Length 422;
Best Local Similarity 50.0%; Pred. No. 2e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wxxxayxg 8
|| | |
Db 5 WRTEIYSG 12

RESULT 4
US-08-486-414-6
Sequence 6, Application US/08486414B
Patent No. 6136318
GENERAL INFORMATION:
APPLICANT: Cochran, Mark D.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUSES AND USES THEREOF
FILE REFERENCE: 42771D
CURRENT APPLICATION NUMBER: US/08/486,414B
CURRENT FILING DATE: 1995-06-07
NUMBER OF SEQ ID NOS: 46
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 6
LENGTH: 422
TYPE: PRT
ORGANISM: Fowlpox virus
US-08-486-414-6

Query Match 75.0%; Score 27; DB 4; Length 422;
Best Local Similarity 50.0%; Pred. No. 2e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wxxxayxg 8
|| | |
Db 5 WRTEIYSG 12

RESULT 5
PCT-US94-01826A-6
Sequence 6, Application PC/TUS9401826A
GENERAL INFORMATION:
APPLICANT: Syntro Corporation, et al.
TITLE OF INVENTION: Recombinant Fowlpox Virus S-FPV-043 and Uses Thereof
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01826A
FILING DATE: 28-FEB-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: White Esq, John P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)977-9550
TELEFAX: (212)664-0525
TELEX: 422523
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 422 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01826A-6

Query Match 75.0%; Score 27; DB 5; Length 422;
Best Local Similarity 50.0%; Pred. No. 2e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 wrxxayxg 8
111111
DB 5 WRTEIYSG 12

RESULT 6
PCT-US94-02252A-6
Sequence 6, Application PC/TUS9402252A
GENERAL INFORMATION:
APPLICANT: Syntro Corporation, et al.
TITLE OF INVENTION: Recombinant Fowlpox viruses and Uses Thereof
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/02252A
FILING DATE: 28-FEB-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: White Esq, John P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)977-9550
TELEFAX: (212)664-0525
TELEX: 422523
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 422 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-02252A-6

Query Match 75.0%; Score 27; DB 5; Length 422;
Best Local Similarity 50.0%; Pred. No. 2e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 wrxxayxg 8
111111
DB 5 WRTEIYSG 12

RESULT 7
US-07-876-280-4
Sequence 4, Application US/07876280
Patent No. 5262158
GENERAL INFORMATION:
APPLICANT: Payne, Jewel M.
APPLICANT: Cannon, Raymond J.C.
APPLICANT: Bagley, Angela L.
TITLE OF INVENTION: No. 5262158e1 Bacillus thuringiensis Isolates for
NUMBER OF SEQUENCES: Controlling Acarides
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/876,280
FILING DATE: 19920430
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: M/S 104
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1289 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: BACILLUS THURINGIENSIS
INDIVIDUAL ISOLATE: PS17
IMMEDIATE SOURCE:
CLONE: E. coli NM522(pMC 1628) NRRL B-18652
US-07-876-280-4

Query Match 75.0%; Score 27; DB 1; Length 1289;
Best Local Similarity 50.0%; Pred. No. 5.6e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 wrxxayxg 8
111111
DB 382 WRAQYGG 389

RESULT 8
US-07-675-772-4
Sequence 4, Application US/07675772
Patent No. 5262399
GENERAL INFORMATION:
APPLICANT: HICKLE, Leslie A.
APPLICANT: SICK, August J.

APPLICANT: Schwab, George E.
APPLICANT: Narva, Kenneth E.
APPLICANT: Payne, Jewel M.
TITLE OF INVENTION: No. 5262399e1 Compositions and Methods for the Control of
TITLE OF INVENTION: Flukes
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROMAN SALIWANCHIK
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/675,772
FILING DATE: 19910327
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SALIWANCHIK, ROMAN
REGISTRATION NUMBER: 21,023
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1289 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: BACILLUS THURINGIENSIS
INDIVIDUAL ISOLATE: PS17
IMMEDIATE SOURCE:
LIBRARY: LAMBDA GEM (TM) - 11 LIBRARY OF KENNETH NARVA
CLONE: 17B
US-07-675-772-4

Query Match 75.0%; Score 27; DB 1; Length 1289;
Best Local Similarity 50.0%; Pred. No. 5.6e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wxxyxxyx 8
DB 382 WRAAOYGG 389

RESULT 9
US-08-063-170-4
Sequence 4, Application US/08063170
Patent No. 5350576
GENERAL INFORMATION:
APPLICANT: Kim, Leo
APPLICANT: Schwab, George E.
TITLE OF INVENTION: Compositions and Methods for Inducing an Immune
TITLE OF INVENTION: Response for Protection Against Endoparasites and Exoparasites
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: DAVID R. SALIWANCHIK
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: IBM PC compatible

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/063,170
FILING DATE: 19930517
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,141
FILING DATE: 30-SEP-1991
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER: US 07/759,248
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: SALIWANCHIK, DAVID R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: M/S 103.C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1289 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus thuringiensis
STRAIN: PS17
INDIVIDUAL ISOLATE: PS17b
US-08-063-170-4

Query Match 75.0%; Score 27; DB 1; Length 1289;
Best Local Similarity 50.0%; Pred. No. 5.6e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wxxyxxyx 8
DB 382 WRAAOYGG 389

RESULT 10
US-08-158-232-4
Sequence 4, Application US/08158232
Patent No. 5596071
GENERAL INFORMATION:
APPLICANT: Payne, Jewel
APPLICANT: Kennedy, M. Keith
APPLICANT: Randall, John Brooks
APPLICANT: Meier, Henry
APPLICANT: Vick, Heidi Jane
APPLICANT: Foncerrada, Luis
APPLICANT: Schnepf, H. Ernest
APPLICANT: Schwab, George E.
TITLE OF INVENTION: No. 5596071e1 Bacillus thuringiensis Toxins Active
TITLE OF INVENTION: Against Hymenopteran Pests
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/158,232
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/887,980
FILING DATE: 22-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/797,645
FILING DATE: 25-NOV-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/703,977
FILING DATE: 22-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1289 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: BACILLUS THURINGIENSIS
STRAIN: PS17
INDIVIDUAL ISOLATE: PS17B
IMMEDIATE SOURCE:
CLONE: E. coli NM522(pMYC1628) NRRL B-18652
US-08-158-232-4

Query Match      75.0%; Score 27; DB 1; Length 1289;
Best Local Similarity 50.0%; Pred. No. 5.6e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 wrxxayxg 8
      11 111
Db      382 WRAQYGG 389

RESULT 11
US-08-304-626-4
Sequence 4, Application US/08304626
Patent No. 5616495
GENERAL INFORMATION:
APPLICANT: Payne, Jewel M.
APPLICANT: Kennedy, M. Keith
APPLICANT: Randall, John Brooks
APPLICANT: Meier, Henry
APPLICANT: Ullick, Heidi Jane
APPLICANT: Foncetrada, Luis
APPLICANT: Schepf, Harry E.
APPLICANT: Schwab, George E.
TITLE OF INVENTION: No. 5616495el Bacillus thuringiensis Isolates
TITLE OF INVENTION: Active Against Hymenopteran Pests and Genes Encoding
TITLE OF INVENTION: Hymanopteran-Active Toxins
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/304,626
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/887,980
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: M/SCJ 104
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1289 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: BACILLUS THURINGIENSIS
STRAIN: PS17
INDIVIDUAL ISOLATE: PS17B
IMMEDIATE SOURCE:
CLONE: E. coli NM522(pMYC1628) NRRL B-18652
US-08-304-626-4

Query Match      75.0%; Score 27; DB 1; Length 1289;
Best Local Similarity 50.0%; Pred. No. 5.6e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 wrxxayxg 8
      11 111
Db      382 WRAQYGG 389

RESULT 12
US-08-316-301A-4
Sequence 4, Application US/08316301A
Patent No. 5753492
GENERAL INFORMATION:
APPLICANT: Schepf, Harry E.
APPLICANT: Schwab, George E.
APPLICANT: Payne, Jewel M.
APPLICANT: Narva, Kenneth E.
APPLICANT: Foncetrada, Luis
TITLE OF INVENTION: No. 5753492el Nematode-Active Toxins and Genes
TITLE OF INVENTION: Which Code Therefor
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/316,301A
```

```

: FILING DATE: 30-SEP-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/871,510
: FILING DATE: 23-APR-1992
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/693,018
: FILING DATE: 03-MAY-1991
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/565,544
: FILING DATE: 10-AUG-1990
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/084,653
: FILING DATE: 12-AUG-1987
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/830,050
: FILING DATE: 31-JAN-1992
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Lloyd, Jeff
: REGISTRATION NUMBER: 35,589
: REFERENCE/DOCKET NUMBER: MA20CCCD1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 904-375-8100
: TELEFAX: 904-372-5800
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1289 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEetical: YES
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: BACILLUS THURINGIENSIS
: INDIVIDUAL ISOLATE: PS17
: IMMEDIATE SOURCE:
: CLONE: E. coli NM522(PMYC 1628) NRRL B-18652
: US-08-316-301A-4

Query Match          75.0%; Score 27; DB 1; Length 1289;
Best Local Similarity 50.0%; Pred. No. 5.6e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      1 wxxyxg 8
      || ||
Db      382 WRAAYGG 389

RESULT 13
US-08-611-928-4
: Sequence 4, Application US/08611928
: Patent No. 5824792
: GENERAL INFORMATION:
: APPLICANT: Payne, Jewel
: APPLICANT: Kennedy, M. Keith
: APPLICANT: Randall, John Brooks
: APPLICANT: Meier, Henry
: APPLICANT: Ulick, Heidi Jane
: APPLICANT: Foncecerra, Luis
: APPLICANT: Schaeff, H. Ernest
: APPLICANT: Schwab, George E.
: APPLICANT: Fu, Jenny
: TITLE OF INVENTION: No. 5824792el Bacillus thuringiensis Toxins Active
: NUMBER OF SEQUENCES: 51
: CORRESPONDENCE ADDRESS:
:
```

```

: ADDRESSEE: David R. Saliwanchik
: STREET: 2421 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: USA
: ZIP: 32606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/611,928
: FILING DATE: 06-MAR-1996
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/158,232
: FILING DATE: 24-NOV-1993
: APPLICATION NUMBER: US 07/887,980
: FILING DATE: 22-MAY-1992
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/797,645
: FILING DATE: 25-NOV-1991
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/703,977
: FILING DATE: 22-MAY-1991
: CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
: NAME: Saliwanchik, David R.
: REGISTRATION NUMBER: 31,794
: REFERENCE/DOCKET NUMBER: M/SCI104.C1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 904-375-8100
: TELEFAX: 904-372-5800
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1289 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEtical: YES
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: BACILLUS THURINGIENSIS
: STRAIN: PS17
: INDIVIDUAL ISOLATE: PS17b
: IMMEDIATE SOURCE:
: CLONE: E. coli NM522(PMYC1628) NRRL B-18652
: US-08-611-928-4

Query Match          75.0%; Score 27; DB 2; Length 1289;
Best Local Similarity 50.0%; Pred. No. 5.6e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      1 wxxyxg 8
      || ||
Db      382 WRAAYGG 389

RESULT 14
US-09-173-891-4
: Sequence 4, Application US/09173891
: Patent No. 6077937
: GENERAL INFORMATION:
: APPLICANT: Payne, Jewel
: APPLICANT: Kennedy, M. Keith
: APPLICANT: Randall, John Brooks
: APPLICANT: Meier, Henry
: APPLICANT: Ulick, Heidi Jane
:
```



```

: APPLICANT: Foncerrada, Luis
: APPLICANT: Schnepf, H. Ernest
: APPLICANT: Schwab, George E.
: APPLICANT: Fu, Jenny
: TITLE OF INVENTION: No. 6077937e1 Bacillus thuringiensis Toxins Active
: TITLE OF INVENTION: Against Hymenopteran Pests
: NUMBER OF SEQUENCES: 51
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: David R. Saliwanchik
: STREET: 2421 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: USA
: ZIP: 32606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/173,891
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/158,232
: FILING DATE:
: APPLICATION NUMBER: US 07/887,980
: FILING DATE: 22-MAY-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/797,645
: FILING DATE: 25-NOV-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/703,977
: FILING DATE: 22-MAY-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Saliwanchik, David R.
: REGISTRATION NUMBER: 31,794
: REFERENCE/DOCKET NUMBER: M/SCJ104.C1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 904-375-8100
: TELEFAX: 904-372-5800
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1289 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: YES
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: BACILLUS THURINGIENSIS
: STRAIN: PS17
: INDIVIDUAL ISOLATE: PS17b
: IMMEDIATE SOURCE:
: CLONE: E. coli NM522(pMVC1628) NRRL B-18652
: US-09-173-891-4

Query Match 75.0%; Score 27; DB 3; Length 1289;
Best Local Similarity 50.0%; Pred. No. 5.6e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 WXXAYXG 8
DB 382 WRAQYCG 389

```

```

: APPLICANT: Schnepf, Harry E.
: APPLICANT: Schwab, George E.
: APPLICANT: Payne, Jewel M.
: APPLICANT: Narva, Kenneth E.
: APPLICANT: Foncerrada, Luis
: TITLE OF INVENTION: No. 6166195e1 Nematode-Active Toxins and Genes which Code
: TITLE OF INVENTION: Therefor
: FILE REFERENCE: MA-20CCCD2
: CURRENT APPLICATION NUMBER: US/09/076,137B
: CURRENT FILING DATE: 1998-05-12
: EARLIER APPLICATION NUMBER: 08/316,301
: EARLIER FILING DATE: 1994-09-30
: NUMBER OF SEQ ID NOS: 42
: SOFTWARE: Patentln Ver. 2.1
: SEQ ID NO 4
: LENGTH: 1289
: TYPE: PRT
: ORGANISM: Bacillus thuringiensis
: US-09-076-137-4

Query Match 75.0%; Score 27; DB 4; Length 1289;
Best Local Similarity 50.0%; Pred. No. 5.6e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 WXXAYXG 8
DB 382 WRAQYCG 389

RESULT 16
PCT-US92-03624-4
: Sequence 4, Application PC/TUS9203624
: GENERAL INFORMATION:
: APPLICANT: Schnepf, Harry E.
: APPLICANT: Schwab, George E.
: APPLICANT: Payne, Jewel M.
: APPLICANT: Narva, Kenneth E.
: APPLICANT: Foncerrada, Luis
: TITLE OF INVENTION: Novel Nematode-Active Toxins and Genes
: TITLE OF INVENTION: Which Code Therefor
: NUMBER OF SEQUENCES: 40
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: David R. Saliwanchik
: STREET: 2421 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: USA
: ZIP: 32606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US92/03624
: FILING DATE: 19920501
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Saliwanchik, David R.
: REGISTRATION NUMBER: 31,794
: REFERENCE/DOCKET NUMBER: MA20C2C1C1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 904-375-8100
: TELEFAX: 904-372-5800
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1289 amino acids
: TYPE: AMINO ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: YES

```

ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: BACILLUS THURINGIENSIS
INDIVIDUAL ISOLATE: PS17
IMMEDIATE SOURCE:
CLONE: E. coli NM522(pMC 1628) NRRL B-18652
PCT-US92-03624-4

Query Match 75.0%; Score 27; DB 5; Length 1289;
Best Local Similarity 50.0%; Pred. No. 5.6e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 WXXXYXG 8
11 11
DB 382 WRAQYCG 389

RESULT 17
US-08-158-232-43
Sequence 43, Application US/08158232
Patent No. 5596071

GENERAL INFORMATION:

APPLICANT: Payne, Jewel
APPLICANT: Kennedy, M. Keith
APPLICANT: Randall, John Brooks
APPLICANT: Meier, Henry
APPLICANT: Uick, Heidi Jane
APPLICANT: Foncerrada, Luis
APPLICANT: Schepf, H. Ernest
APPLICANT: Schwab, George E.
APPLICANT: Fu, Jenny
TITLE OF INVENTION: No. 5596071el Bacillus thuringiensis Toxins Active
TITLE OF INVENTION: Against Hymenopteran Pests
NUMBER OF SEQUENCES: 51

CORRESPONDENCE ADDRESS:

ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL

COUNTRY: USA
ZIP: 32606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/158, 232

FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/887, 980

FILING DATE: 22-MAY-1992

APPLICATION NUMBER: US 07/797, 645

FILING DATE: 25-NOV-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/703, 977

FILING DATE: 22-MAY-1991

ATTORNEY/AGENT INFORMATION:

NAME: Saliwanchik, David R.

REGISTRATION NUMBER: 31,794

REFERENCE/DOCKET NUMBER: M/SCJ104.C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 904-375-8100

TELEFAX: 904-372-5600

INFORMATION FOR SEQ ID NO: 43:

SEQUENCE CHARACTERISTICS:

LENGTH: 1220 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:

ORGANISM: Bacillus thuringiensis
INDIVIDUAL ISOLATE: 86Q3

IMMEDIATE SOURCE:

LIBRARY: LambdaGem (TM) - 11 library of Luis

CLONE: 86Q3c

US-08-158-232-43

Query Match 72.2%; Score 26; DB 1; Length 1220;
Best Local Similarity 50.0%; Pred. No. 8.5e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 WXXXYXG 8
11 11
DB 382 WRAQYCG 389

RESULT 18
US-08-611-928-43
Sequence 43, Application US/08611928
Patent No. 5824792

GENERAL INFORMATION:

APPLICANT: Payne, Jewel
APPLICANT: Kennedy, M. Keith
APPLICANT: Randall, John Brooks
APPLICANT: Meier, Henry
APPLICANT: Uick, Heidi Jane
APPLICANT: Foncerrada, Luis
APPLICANT: Schepf, H. Ernest
APPLICANT: Schwab, George E.
APPLICANT: Fu, Jenny

TITLE OF INVENTION: No. 5824792el Bacillus thuringiensis Toxins Active
TITLE OF INVENTION: Against Hymenopteran Pests
NUMBER OF SEQUENCES: 51

CORRESPONDENCE ADDRESS:

ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL

COUNTRY: USA
ZIP: 32606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/611, 928

FILING DATE: 06-MAR-1996

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/158, 232

FILING DATE: 24-NOV-1993

APPLICATION NUMBER: US 07/887, 980

FILING DATE: 22-MAY-1992

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/797, 645

FILING DATE: 25-NOV-1991

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/703, 977

FILING DATE: 22-MAY-1991

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Saliwanchik, David R.

REGISTRATION NUMBER: 31,794

REFERENCE/DOCKET NUMBER: M/SCJ104.C1

TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 1220 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus thuringiensis
INDIVIDUAL ISOLATE: 86Q3
IMMEDIATE SOURCE:
LIBRARY: Lambdagem (TM) - 11 library of Luis
LIBRARY: Foncerada
CLONE: 86Q3c
US-08-611-928-43

Query Match 72.2% Score 26: DB 2: Length 1220:
Best Local Similarity 50.0% Pred. No. 8.5e+02:
Matches 4: Conservative 0: Mismatches 4: Indels 0: Gaps 0:

QY 1 wrxxayxg 8
|| | |
Db 382 WRAGQYGC 389

RESULT 19
US-09-173-891-43
Sequence 43, Application US/09173891
Patent No. 6077937
GENERAL INFORMATION:
APPLICANT: Payne, Jewel
APPLICANT: Kennedy, M. Keith
APPLICANT: Randall, John Brooks
APPLICANT: Meier, Henry
APPLICANT: Ulick, Heidi Jane
APPLICANT: Foncerada, Luis
APPLICANT: Schnepf, H. Ernest
APPLICANT: Schwab, George E.
APPLICANT: Fu, Jenny
TITLE OF INVENTION: No. 6077937e1 Bacillus thuringiensis Toxins Active
TITLE OF INVENTION: Against Hymenopteran Pests
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/173,891
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/158,232
FILING DATE:
APPLICATION NUMBER: US 07/887,980
FILING DATE: 22-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/797,645
FILING DATE: 25-NOV-1991
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/703,977
FILING DATE: 22-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: M/SCJ104.C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 1220 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus thuringiensis
INDIVIDUAL ISOLATE: 86Q3
IMMEDIATE SOURCE:
LIBRARY: Lambdagem (TM) - 11 library of Luis
LIBRARY: Foncerada
CLONE: 86Q3c
US-09-173-891-43

Query Match 72.2% Score 26: DB 3: Length 1220:
Best Local Similarity 50.0% Pred. No. 8.5e+02:
Matches 4: Conservative 0: Mismatches 4: Indels 0: Gaps 0:

QY 1 wrxxayxg 8
|| | |
Db 382 WRAGQYGC 389

RESULT 20
5281530-3
Patent No. 5281530
APPLICANT: SICK, AUGUST J.; SCHWAB, GEORGE E.; PAYNE,
JEWEL M.
TITLE OF INVENTION: GENES ENCODING NEMATODE-ACTIVE TOXINS
CLONED FROM BACILLUS
NUMBER OF SEQUENCES: 4
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/557,246
FILING DATE: 24-JUL-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 535,810
FILING DATE: 11-JUN-1990
APPLICATION NUMBER: 84,653
FILING DATE: 12-AUG-1987
SEQ ID NO: 3:
LENGTH: 1289
5281530-3

Query Match 72.2% Score 26: DB 6: Length 1289:
Best Local Similarity 50.0% Pred. No. 8.9e+02:
Matches 4: Conservative 0: Mismatches 4: Indels 0: Gaps 0:

QY 1 wrxxayxg 8
|| | |
Db 382 WRAGQYGC 389

Search completed: January 14, 2002, 07:23:44
Job time: 75 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:37:30 ; Search time 63.57 Seconds
(without alignments)
9.586 Million cell updates/sec

Title: 09-185908-1g
Perfect score: 35
Sequence: 1 wrxxsfxxg 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database :
1: pIR_68: *
2: pIR2: *
3: pIR3: *
4: pIR4: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	88.6	206	2	H75258
2	30	85.7	1083	2	S54293
3	29	82.9	199	1	TJBPXL
4	29	82.9	224	2	B85584
5	29	82.9	777	2	T00208
6	28	82.9	1715	2	G84429
7	28	80.0	70	2	H64429
8	28	80.0	308	2	S22923
9	28	80.0	528	2	G70854
10	28	80.0	528	2	T45418
11	28	80.0	1180	1	NCEKX5
12	28	80.0	1180	2	G85933
13	28	80.0	1208	2	B82091
14	27	77.1	159	2	S41178
15	27	77.1	212	2	S72873
16	27	77.1	278	2	H71119
17	27	77.1	280	2	A39484
18	27	77.1	339	2	A84529
19	27	77.1	330	2	T34972
20	27	77.1	344	2	T34153
21	27	77.1	370	2	B83191
22	27	77.1	551	2	T16557
23	27	77.1	998	2	G83022
24	26	74.3	85	2	G70659
25	26	74.3	122	2	F82861
26	26	74.3	141	2	E72594
27	26	74.3	154	2	E70971
28	26	74.3	198	2	G85506
29	26	74.3	231	2	T50853

30	26	74.3	239	2	S65825	hypothetical prote
31	26	74.3	259	2	T48851	response regulator
32	26	74.3	261	2	C64948	probable membrane
33	26	74.3	261	2	E85798	hypothetical prote
34	26	74.3	346	2	A82971	low specificity 1-
35	26	74.3	362	2	JU0353	hypothetical 39.7k
36	26	74.3	371	2	A71359	conserved hypotet
37	26	74.3	379	2	D70786	probable gcvT prot
38	26	74.3	392	2	D70475	conserved hypotet
39	26	74.3	404	2	G83322	hypothetical prote
40	26	74.3	409	1	S32905	serine proteinase
41	26	74.3	458	2	T16123	hypothetical prote
42	26	74.3	483	2	G84113	hypothetical prote
43	26	74.3	604	2	SC0182	hypothetical prote
44	26	74.3	888	2	A71720	ATP-binding transp
45	26	74.3	937	2	S78561	hypothetical prote
46	26	74.3	2082	2	T37056	CS3 pilin synthet
47	26	74.3	2154	2	A84669	probable multi-dom
48	26	74.3	2347	1	TYHURS	hypothetical prote
49	25	71.4	54	2	F86583	kinase-related pro
50	25	71.4	54	2	G72040	hypothetical prote

ALIGNMENTS

RESULT 1
H75258
probable 3-demethylubiquitinone-9-3-methyltransferase - Deinococcus radiodurans (strain
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: H75258
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: H75258
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-206 <WH>
A:Cross-references: GB:AE002085; GB:AE00513; NID:g6460383; PIDN:AAF12100.1; PID:g646
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR2562
A:Map position: 1

Query Match 88.6%; Score 31; DB 2; Length 206;
Best Local Similarity 62.5%; Pred. No. 6.6;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxsfxxg 8
DB 111 WRJASFDFG 118

RESULT 2
S54293
regulator protein p122-RhoGAP - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 01-Sep-1995
C:Accession: S54293
R:Homma, Y.; Emori, Y.
EMBO J. 14, 286-291, 1995
A:Title: A dual functional signal mediator showing RhoGAP and phospholipase C-delta s
A:Reference number: S54293; MUID:95137008
A:Accession: S54293
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1083 <HW>
A:Cross-references: EMBL:D31962

Query Match 85.7% Score 30; DB 2; Length 1083;
Best Local Similarity 62.5%; Pred. No. 55;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxsxfxg 8
|||
DB 393 WRGGSFHG 400

RESULT 3
TUBPKL
tail assembly protein K - phage lambda
C:Species: phage lambda
C:Date: 13-Jun-1983 #sequence_revision 13-Jun-1983 #text_change 23-Jul-1999
A:Accession: H43009; G43013; A04355
R:Daniels, D.
submitted to the Nucleic Acid Sequence Database, September 1982
A:Reference number: A94614
A:Accession: H43009
A:Molecule type: DNA
A:Residues: 1-199 <DAN>
R:Sanger, F.; Coulson, A.R.; Hong, G.F.; Hill, D.F.; Petersen, G.B.
J. Mol. Biol. 162, 729-773, 1982
A:Title: Nucleotide sequence of bacteriophage lambda DNA.
A:Reference number: A92891; MUID:83189071
A:Accession: G43013
A:Molecule type: DNA
A:Residues: 1-199 <SAN>
A:Cross-references: GB:002459; GB:M17233; GB:M24325; GB:V00636; GB:X00906; NID:g215104;
A:Note: there are two possible initiation sites for gene K translation, the codon for 1-
C:Comment: Gene K protein is involved in the assembly of the initiator complex for tail
C:Genetics:
A:Gene: K
A:Map position: 29.43-30.66
C:Superfamily: phage lambda tail assembly protein K

Query Match 82.9% Score 29; DB 1; Length 199;
Best Local Similarity 50.0%; Pred. No. 18;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxsxfxg 8
|||
DB 180 WRASAFHG 187

RESULT 4
B85584
probable tail component of prophage CP-933K Z0978 [imported] - Escherichia coli (strain
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
A:Accession: B85584
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B85584
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-224 <STO>
A:Cross-references: GB:AE005174; NID:g12513746; PIDN:AA55134.1; GSPDB:GN00145; UMGPR:Z09
C:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z0978
C:Superfamily: phage lambda tail assembly protein K

Query Match 82.9% Score 29; DB 2; Length 224;
Best Local Similarity 50.0%; Pred. No. 21;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxsxfxg 8
|||
DB 205 WRASAFHG 212

RESULT 5
T00208
transposase-like protein - fungus (Fusarium oxysporum) Ac-type transposon Tfo1
C:Species: Fusarium oxysporum
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
A:Accession: T00208
R:Okuda, M.; Ikeda, K.; Namiki, F.; Nishi, K.; Tsuge, T.
Mol. Gen. Genet. 258, 599-607, 1998
A:Title: Tfo1: an Ac-like transposon from the plant pathogenic fungus Fusarium oxyspo
A:Reference number: 214125; MUID:98334103
A:Accession: T00208
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-777 <OKU>
A:Cross-references: EMBL:AB008746; NID:g3410895; PIDN:BA32244.1; PTD:g3410896
A:Experimental source: strain MAF305118
C:Genetics:
A:Mobile element: Ac-type transposon Tfo1

Query Match 82.9% Score 29; DB 2; Length 777;
Best Local Similarity 62.5%; Pred. No. 68;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxsxfxg 8
|||
DB 402 WRKGSFHG 409

RESULT 6
G84429
hypotheical protein At2g01840 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
A:Accession: G84429
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.T.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: G84429
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1715 <STO>
A:Cross-references: GB:AE002093; NID:g4522005; PIDN:AAD21778.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g01840
A:Map position: 2

Query Match 82.9% Score 29; DB 2; Length 1715;
Best Local Similarity 50.0%; Pred. No. 15e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxsxfxg 8
|||
DB 1450 WRSANFSG 1457

RESULT 7
H64449
ribosomal protein L24E - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
A:Accession: H64449
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak

; Relch, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
; Science, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
; tson, J.D., 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C
A:Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*
A:Reference number: A64300; MUID:96337999
A:Accession: H64449
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-70 <BU>
A:Cross-references: GB:U67562; GB:L77117; NID:92826374; PIDN:AB99205.1; PID:g1592325; T
C:Genetics:
A:Map position: REV1146494-1146282
C:Superfamily: Halorarcula ribosomal protein HU21

Query Match 80.0%; Score 28; DB 2; Length 70;
Best Local Similarity 62.5%; Pred. No. 11;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfxxg 8
|||
Db 4 WRTCSFSG 11

RESULT 8
S22923
ubiquinol--cytochrome-c reductase (EC 1.10.2.2) cytochrome b - Andean flicker mitochondr
C:Species: mitochondrion Colaptes auratus ruficollis (Andean flicker)
C:Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 04-Mar-2000
C:Accession: S22923
R:Edwards, S.V.; Arcander, P.; Wilson, A.C.
Proc. R. Soc. Lond. B Biol. Sci. 243, 99-107, 1991
A:Title: Mitochondrial resolution of a deep branch in the genealogical tree for perching
A:Reference number: S22919; MUID:91288587
A:Accession: S22923
A:Status: translation not shown
A:Residues: 1-308 <EDU>
A:Molecule type: DNA
A:Cross-references: EMBL:X60949; NID:g12892; PIDN:CAA3284.1; PID:g12893
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC1
C:Superfamily: cytochrome b; cytochrome b homology; cytochrome b6 homology; plastocinon
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; mitochondrion;
F:1-307/Domain: cytochrome b homology (fragment) <CBH>
F:1-177/Domain: cytochrome b6 homology (fragment) <CB6>
F:4-20/Domain: transmembrane #status predicted <TM1>
F:49-67/Domain: transmembrane #status predicted <TM2>
F:85-101/Domain: transmembrane #status predicted <TM3>
F:146-168/Domain: transmembrane #status predicted <TM4>
F:189-307/Domain: plastocinon--plastocyanin reductase 17k protein homology <17k>
F:189-213/Domain: transmembrane #status predicted <TM5>
F:256-272/Domain: transmembrane #status predicted <TM6>
F:65,164/Binding site: heme iron (His) (axial ligands) (high potential) #status predicte
F:150/Binding site: heme iron (His) (axial ligand) (low potential) #status predicted

Query Match 80.0%; Score 28; DB 2; Length 308;
Best Local Similarity 62.5%; Pred. No. 47;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfxxg 8
|||
Db 103 WRKMSFSG 110

RESULT 9
S22923
probable *serA* protein - *Mycobacterium tuberculosis* (strain H37RV)
C:Species: *Mycobacterium tuberculosis*
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: G70854

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: G70854
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-528 <COL>
A:Cross-references: GB:AL021287; GB:AL123456; NID:g3261508; PIDN:CAA16081.1; PID:g279
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: *serA*
C:Superfamily: *Bacillus* phosphoglycerate dehydrogenase

Query Match 80.0%; Score 28; DB 2; Length 528;
Best Local Similarity 50.0%; Pred. No. 79;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfxxg 8
|||
Db 129 WKSSFSFG 136

RESULT 10
T45418
phosphoglycerate dehydrogenase [imported] - *Mycobacterium leprae*
C:Species: *Mycobacterium leprae*
C:Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 18-Feb-2000
C:Accession: T45418
R:Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1997
A:Reference number: Z16918
A:Accession: T45418
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-528 <PAR>
A:Cross-references: EMBL:Z99263; PIDN:CAB16440.1
A:Experimental source: cosmid B637
C:Genetics:
A:Note: *serA*
C:Superfamily: *Bacillus* phosphoglycerate dehydrogenase

Query Match 80.0%; Score 28; DB 2; Length 528;
Best Local Similarity 50.0%; Pred. No. 79;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfxxg 8
|||
Db 129 WKSSFSFG 136

RESULT 11
NC0035
exodeoxyribonuclease V (EC 3.1.11.5) 135k chain - *Escherichia coli*
N:Alternate names: exonuclease 135k polypeptide; recBC DNase 135k polypeptide
C:Species: *Escherichia coli*
C:Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 19-Jan-2001
C:Accession: A25532; E65064
R:Finch, P.W.; Storey, A.; Chapman, K.E.; Brown, K.; Hickson, I.D.; Emmerson, P.T.
Nucleic Acids Res. 14, 8573-8582, 1986
A:Title: Complete nucleotide sequence of the *Escherichia coli* recB gene.
A:Reference number: A25532; MUID:87066729
A:Accession: A25532
A:Molecule type: DNA
A:Residues: 1-1180 <FIN>
A:Cross-references: GB:X04581; NID:g42680; PIDN:CAA28250.1; PID:g42682
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277 1453-1462, 1997
 A:Title: The complete genome sequence of *Escherichia coli* K-12.
 A:Reference number: A64720; MUID:97426617
 A:Accession: E65064
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1180 <BLAT>
 A:Cross-references: GB:AE000365; GB:U000096; NID:92367165; PIDN:AAC75859.1; PID:g1789183
 A:Experimental source: strain K-12, substrain MG1655
 C:Comment: This enzyme is required for efficient DNA repair; it catalyzes the unwinding
 11 of these activities require concomitant hydrolysis of ATP.
 C:Genetics:
 A:Gene: recB
 A:Map position: 61 min
 C:Superfamily: exodeoxyribonuclease V 135K chain
 C:Keywords: ATP, DNA repair, hydrolyase; nucleotide binding; P-loop
 [23-30/Region: nucleotide-binding motif A (P-loop)]

Query Match	80.0%	Score 28;	DB 1;	Length 1180;
Best Local Similarity	50.0%;	Pred. No. 1.7e+02;		
Matches 4;	Conservative 1;	Mismatches 3;	Indels 0;	Gaps 0;
OY 1	wxxsfxg 8			
11	1: 1			
901	WRVTSYSG 908			

RESULT 12
G85933
hypothetical protein recB [imported] - *Escherichia coli* (strain O157:H7)
C:Species: *Escherichia coli*
C:Date: 16-Feb-2001 #sequence-revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: G85933
C:Penna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
Hiller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoustis, K.; Apodaca,
Nature 409, 529-533, 2001
A:title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: G85933
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1180 <STO>
A:Cross-references: GB:AE005174; MID:g12517302; PIDN:AAC57931.1; GSPDB:GN00145; UWGP:Z44
A:Experimental source: strain O157:H7, Substrain EDL933
C:Genetics:
A:Gene: recB
C:Superfamily: exodeoxyribonuclease V 135k chain

```

Query Match          80.0%; Score 28; DB 2; Length 1180;
Best Local Similarity 50.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY      1 WRXXSfXg 8
          ||| : |
Db       901 WRVTSYSG 908

RESULT 13
B82091
exodeoxyribonuclease V, 135 kDa chain VC2320 [Imported] - Vibrio cholerae (Strain N1696)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence-revision 20-Aug-2000 #text-change 02-Feb-2001
C:Accession: B82091
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
  Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoli, I.; Sellers, F.
  1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
  Nature 406 477-483, 2000
A:Title: DNA Sequence of Both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: B82091
A:Status: preliminary

```

A:Molecule type: DNA
A:Residues: 1-1208 <HEI>
A:Cross-references: GB:AE004303; GB:AE003852; NID:99656890; PIDD:AAF95464.1; GSPDB:GM
A:Experimental source: serogroup O1; strain N16961; bioType E1 Tor
C:Genetics:
A:Gene: VC2320
A:Map position: 1
C:Superfamily: exodeoxyribonuclease V 135k chain

```

Query Match      80.0%;  Score 28;  DB 2;  Length 1208;
Best Local Similarity 50.0%;  Pred. No. 1.8e+02;
Matches 4;  Conservative 1;  Mismatches 3;  Indels 0;  Gaps 0;

QY      1 wxxxxfixg 8
        |||:|
Db      906 WRVTSYSG 913

```

```

RESULT 14
S41178
gene 36 protein - phage SP1
C:Species: phage SP1
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 11-May-2000
C:Accession: S43804; T42337; S41178
R:Pedre, X.; Weisse, F.; Chai, S.; Lueder, G.; Alonso, J.C.
J. Mol. Biol. 236, 1324-1340, 1994
A:Title: Analyls of cis and trans acting elements required for the initiation of DNA
A:Reference number: S43798; M0ID:94172631
A:Accession: S43804
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-159 <PE2>
A:Cross-references: EMBL:X67865; NID:9472886; PID:CAA48055.1; PID:9439635
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, July 1992
R:Alonso, J.C.; Luder, G.; Stieger, A.C.; Chai, S.; Weisse, F.; Trautner, T.A.
Gene 204, 201-212, 1997
A:Title: The complete nucleotide sequence and functional organization of Bacillus sub
A:Reference number: Z22137; M0ID:98094274
A:Accession: T42337
A:Status: Preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-159 <ALO>
A:Cross-references: EMBL:X97918; PID:CAA66491.1
C:Superfamily: single-stranded DNA-binding protein; single-stranded DNA-binding prote
:17-93/Domain: single-stranded DNA-binding protein homology <SSD>

```

```

Query Match      77.1%   Score 27;   DB 2;   Length 159;
Best Local Similarity 50.0%;   Pred. No. 42;
Matches      4;   Conservative      1;   Mismatches      3;   Indels      0;   Gaps
QY      1  WRXSFXX 8
      || : ||
Db      77  WRTGAFEG 84

RESULT 15
S72873
hypothetical protein B2126_F2-70 - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Mar-2001
C:Accession: S72873
R:Smith, D. R.; Robison, K.
submitted to the EMBL Data Library, November 1993
A:Description: Mycobacterium leprae cosmid B2126.
A:Reference number: S72585
A:Accession: S72873
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-212 <SM>
A:Cross-references: EMBL:U00017; NID:g466994; PIDN:AAA17213.1; PID:g467028
A:Genetics:

```


Start codon: CTC

Query Match 77.1%; Score 27; DB 2; Length 212;
Best Local Similarity 50.0%; Pred. No. 56;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxxsfxxg 8
|||
Db 153 WRTIOFSG 160

RESULT 16

H7119
hypothetical protein PH0727 - *Pyrococcus horikoshii*

C:Species: *Pyrococcus horikoshii*
C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 05-Nov-1999

C:Accession: H7119
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekiri, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Oguchi, DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaeon.
A:Reference number: A11000; MUID:98344137

A:Accession: H7119
A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-278 <KAM>
A:Cross-references: GB:AP000003; NID:g3236130; PIDN:BAA29818.1; PID:d1030761; PID:g32571

A:Experimental source: strain OT3
A:Note: This accession replaces an interim accession for a sequence replaced by GenBank

C:Genetics:
A:Gene: PH0727

Query Match 77.1%; Score 27; DB 2; Length 278;
Best Local Similarity 50.0%; Pred. No. 73;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfxxg 8
|||
Db 99 WRVSSYMG 106

RESULT 17

A39484
androgen-withdrawal apoptosis protein RVPI, prostatic - rat

C:Species: *Rattus norvegicus* (Norway rat)
C>Date: 03-Aug-1992 #sequence_revision 03-Aug-1992 #text_change 04-Mar-2000

C:Accession: A39484

R:Brichl, M.M.; Miesfeld, R.L.
Mol. Endocrinol. 5, 1381-1388, 1991

A:Title: Isolation and characterization of transcripts induced by androgen withdrawal in rat L cells.
A:Reference number: A39484; MUID:92130967

A:Accession: A39484
A:Molecule type: mRNA

A:Residues: 1-280 <BR1>
A:Cross-references: GB:M74067; NID:g205857; PIDN:AAA41760.1; PID:g205858

C:Genetics:
A:Gene: RVP.1
C:Superfamily: rat androgen-withdrawal apoptosis protein RVPI

Query Match 77.1%; Score 27; DB 2; Length 280;
Best Local Similarity 50.0%; Pred. No. 73;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfxxg 8
|||
Db 29 WRVSAPFG 36

RESULT 18

A84529

hypothetical protein At2g15440 [imported] - *Arabidopsis thaliana*

C:Species: *Arabidopsis thaliana* (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: A84529

R:Lin, X.; Kuhl, S.; Rounsley, S.D.; Shee, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffatt, K.S.; Cronin, L.A.; Shen, M.; Vankken, S.E.; Umayam, L.; Tallon, euss, D.; Nierman, W.C.; White, O.; Elsen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter

Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.

A:Reference number: A84420; MUID:20083487

A:Accession: A84529
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-329 <STO>

A:Cross-references: GB:AE002093; NID:g4544369; PIDN:AA22280.1; GSPDB:GN00139

C:Genetics:
A:Gene: At2g15440
A:Map position: 2

Query Match 77.1%; Score 27; DB 2; Length 329;
Best Local Similarity 50.0%; Pred. No. 85;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfxxg 8
|||
Db 128 WRSINFGG 135

RESULT 19

T34972
probable membrane protein - *Streptomyces coelicolor*

C:Species: *Streptomyces coelicolor*
C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999

C:Accession: T34972

R:Saunders, D.C.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M
submitted to the EMBL Data Library, August 1999

A:Reference number: Z21563
A:Accession: T34972

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-330 <SAU>
A:Cross-references: EMBL:AL109663; PIDN:CAB52011.1; GSPDB:GN00070; SCOEDB:SC4A10.35C

A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC4A10.35C

Query Match 77.1%; Score 27; DB 2; Length 330;
Best Local Similarity 50.0%; Pred. No. 86;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfxxg 8
|||
Db 31 WRLAFAFG 38

RESULT 20

T34153
hypothetical protein C33H5.1 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T34153

R:Bradshaw, H.; Stellyes, L.
submitted to the EMBL Data Library, November 1995

A:Description: The sequence of C. elegans cosmid C33H5.
A:Reference number: Z21482

A:Accession: T34153
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-344 <BRA>
A:Cross-references: EMBL:U41007; PIDN:AAA82262.1; CESP:C33H5.1

A:Experimental source: strain Bristol N2

C:Genetics:
 A:Gene: CESP:C33H5.1
 A:Introns: 59/3; 123/3; 171/1; 293/3

Query Match 77.18; Score 27; DB 2; Length 344;
 Best Local Similarity 50.0%; Pred. No. 89;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxsfxg 8
 ||:|
 Db 125 WRNTNFTG 132

Search completed: January 14, 2002, 07:37:32
 Job time: 902 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:40:40 ; Search time 37.71 Seconds
(without alignments)
7.778 Million cell updates/sec

Title: 09-185908-1g

Perfect score: 35

Sequence: 1 wxxxxfxg 8

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues 100059

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	88.6	452	1	SPS2_MOUSE
2	29	82.9	199	1	VTAK_LAMB
3	28	80.0	70	1	R2AE_METJA
4	28	80.0	230	1	CLD2_MOUSE
5	28	80.0	308	1	CYB_COLRU
6	28	80.0	528	1	SERA_MYCLE
7	28	80.0	528	1	SERA_MYCTU
8	28	80.0	1180	1	EXS5_ECOLI
9	27	77.1	209	1	CLD4_CERAE
10	27	77.1	209	1	CLD4_HUMAN
11	27	77.1	209	1	CLD4_BRARE
12	27	77.1	210	1	CLD4_MOUSE
13	27	77.1	211	1	CLD1_HUMAN
14	27	77.1	219	1	CLD3_MOUSE
15	27	77.1	219	1	CLD3_MOUSE
16	27	77.1	220	1	CLD3_MOUSE
17	27	77.1	224	1	CLD4_HUMAN
18	27	77.1	448	1	SPS2_HUMAN
19	27	77.1	466	1	ADRO_DROME
20	27	77.1	551	1	YV03_CAEEL
21	26	74.3	159	1	VAT1_CANYP
22	26	74.3	261	1	ZNUB_ECOLI
23	26	74.3	367	1	GCSF_MYCTU
24	26	74.3	409	1	ALP_TRIHA
25	26	74.3	937	1	CS32_ECOLI
26	26	74.3	2347	1	KROS_HUMAN
27	25	71.4	34	1	MYTA_MYRED
28	25	71.4	193	1	CERB_HUMAN
29	25	71.4	193	1	CERB_MOUSE
30	25	71.4	201	1	OM25_BRUV
31	25	71.4	209	1	YC02_KLEPN
32	25	71.4	213	1	OM25_BRUAB
33	25	71.4	213	1	OM25_BRUCA

ALIGNMENTS

RESULT	ID	SPS2_MOUSE	STANDARD	PRT	452 AA.
AC	P97364	15-JUL-1999 (Rel. 38, Created)			
DT	15-JUL-1999 (Rel. 38, Last sequence update)				
DT	30-MAY-2000 (Rel. 39, Last annotation update)				
DE	SELENIIDE, WATER DIKINASE 2 (EC 2.7.9.3) (SELENOPHOSPHATE SYNTHETASE 2)				
DE	(SELENIUM DONOR PROTEIN 2).				
GN	SPS2.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=96017645; PubMed=7588067;				
RA	Guimaraes M.J., Bazan J.F., Zlotnik A., Wiles M.V., Grimaldi J.C.,				
RA	Lee F., McLanahan T.;				
RT	"A new approach to the study of haematopoietic development in the				
RT	yolk sac and embryoid bodies";				
RL	Development 121:3335-3346(1995).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=97140286; PubMed=8986768;				
RA	Guimaraes M.J., Peterson D., Vicari A., Cocks B.G., Copeland N.G.,				
RA	Gilbert D.J., Jenkins N.A., Ferrick D.A., Kastelein R., Bazan J.F.,				
RA	Zlotnik A.;				
RT	"Identification of a novel self homolog from eukaryotes, bacteria,				
RT	and archaea: is there an autoregulatory mechanism in selenocysteine				
RT	metabolism?";				
RL	Proc. Natl. Acad. Sci. U.S.A. 93:15086-15091(1996).				
CC	- FUNCTION: SYNTHESIZES SELENOPHOSPHATE FROM SELENIIDE AND ATP.				
CC	- CATALYTIC ACTIVITY: ATP + SELENIIDE + H(2O) = AMP + SELENOHOSPHATE				
CC	+ PHOSPHATE.				
CC	- COFACTOR: SELENOCYSTEINE. THE ACTIVE-SITE SELENOCYSTEINE IS				
CC	ENCODED BY THE OPAL CODON, UGA.				
CC	- SIMILARITY: BELONGS TO THE SELENOHOSPHATE SYNTHETASE 1 FAMILY.				
CC	CLASS I SUBFAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL: U43285; AAC53024.1; .				
DR	MED: 108388; SPS2.				
DR	InterPro: IPR000728; ATRS-related.				

```

DR Pfam; PF00586; AIRS; 1.
KW Transferrase; Selenium; Selenocysteine; ATP-binding.
FT ACR SITE 63 63 POTENTIAL.
FT SE CYS 63 63
FT SITE 66 66 IMPORTANT FOR CATALYTIC ACTIVITY (BY
FT NP_BIND 322 328 SIMILARITY).
FT DOMAIN 2 9 ATP (POTENTIAL).
FT DOMAIN 433 440 POLY-ALA.
SQ SEQUENCE 452 AA; 47786 MW; 9DA6F7250CFE80EA CRC64;

Query Match
Best Local Similarity 88.6%; Score 31; DB 1; Length 452;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxsfxg 8
DB 52 WRLTFSFG 59

RESULT 2
VTAK_LAMB
AC P03729; STANDARD; PRT; 199 AA.
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE TAIL ASSEMBLY PROTEIN K.
GN K.
OS Bacteriophage lambda.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;
OC Lambda phage group.
OX NCBI_TaxID=10710;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83189071; PubMed=6221115;
RA Sanger F., Coulson A.R., Hong G.F., Hill D.F., Petersen G.B.;
RT "Nucleotide sequence of bacteriophage lambda DNA.";
RL J. Mol. Biol. 162:729-773(1982)
CC -1- FUNCTION: GENE K PROTEIN IS INVOLVED IN THE ASSEMBLY OF THE
INITIATOR COMPLEX FOR TAIL POLYMERIZATION. IT HAS NOT BEEN FOUND
IN THE MATURE PHAGE.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-11 IS THE INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J02459; AAA96551.1; -.
DR PIR; A04355; TJBPKL.
DR InterPro; IPR000064; NLP_C_P60.
DR Pfam; PF00877; NLP_C_P60; 1.
SQ SEQUENCE 199 AA; 23011 MW; CEEB88F01E31ABAE CRC64;

Query Match
Best Local Similarity 82.9%; Score 29; DB 1; Length 199;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxsfxg 8
DB 180 WRASAFSG 187

RESULT 3
R2AE_METJA STANDARD; PRT; 70 AA.
ID R2AE_METJA
AC P54064;

```

```

DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE 50S RIBOSOMAL PROTEIN L24E.
GN RPL24E OR MJ1201.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.L.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii.";
RL Science 273:1058-1073(1996).
CC -1- SIMILARITY: BELONGS TO THE L24E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67562; AAB99205.1; -.
DR TIGR; MJ1201; -.
DR InterPro; IPR000988; Ribosomal_L24E.
DR Pfam; PF01246; Ribosomal_L24E; 1.
DR PROSITE; PS01073; RIBOSOMAL_L24E; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 70 AA; 8249 MW; 3EB6DE18F26E6FCF CRC64;

Query Match
Best Local Similarity 80.0%; Score 28; DB 1; Length 70;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxsfxg 8
DB 4 WRTCSEFG 11

RESULT 4
CID2_MOUSE
ID CID2_MOUSE STANDARD; PRT; 230 AA.
AC O88552;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-2.
GN CLDN2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98311639; PubMed=9647647;
RA Furuse M., Fujita K., Hiragi T., Fujimoto K., Tsukita S.;
RT "Claudin-1 and -2: novel integral membrane proteins localizing at
RT tight junctions with no sequence similarity to occludin.";
RL J. Cell Biol. 141:1539-1550(1998).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.

```

```
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

EMBL: AF072128: AAC27079.1; -.
MCD: MCI:1276110: Clon2.
InterPro: IPR001832: Claudin.
DR InterPro: IPR00729: PMP22_Claudin.
DR Pfam: PF00822: PMP22_Claudin: 1.
DR PRINTS: PRO1077: CLAUDIN.
DR PROSITE: PS01346: CLAUDIN: 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 163 183 POTENTIAL.
SQ SEQUENCE 230 AA: 24483 MW: 38A7C074A1E0D5D2 CRC64:

Query Match Best Local Similarity 80.0%; Score 28; DB 1; Length 230;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfxx 8
|||
Db 30 WRTSSYVG 37

RESULT 5
ID CYB_COLURU STANDARD: PRT: 308 AA.
AC P29635;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CYTOCHROME B (FRAGMENT).
DE MYCYB OR COB OR CYTB.
OS Colaptes rupicola (Audean flicker).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauromorpha; Aves; Neognathae; Piciformes; Picidae; Colaptes.
OX NCBI_TaxID=9222;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-91288587; Pubmed-1676522;
RA Edwards S.V., Arcander P., Wilson A.C.;
RT "Mitochondrial resolution of a deep branch in the genealogical tree
RT for perching birds."
RL Proc. R. Soc. Lond., B, Biol. Sci. 243:99-107(1991).
CC -1- FUNCTION: COMPONENT OF THE UBIQUITIN-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS.
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN.
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RISKE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

```
CC EMBL: X60949: CAA43284.1; -.
CC PIR: S22923; S22923.
CC InterPro: IPR000179: Cyt_b_b6.
CC Pfam: PF00032: cytochrome_b_c1.
CC Pfam: PF00033: cytochrome_b_n: 1.
CC PROSITE: PS00192: CYTOCHROME_B_HEME; FALSE_NEG.
CC PROSITE: PS00193: CYTOCHROME_B_OO: 1.
CC Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
CC Heme.
KW Heme.
FT NON_TER 1 1
FT METAL 51 51 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL 65 65 IRON 2 (HEME B566 AXIAL LIGAND).
FT METAL 150 150 IRON 2 (HEME B566 AXIAL LIGAND).
FT METAL 164 164 IRON 1 (HEME B566 AXIAL LIGAND).
FT NON_TER 308 308
SQ SEQUENCE 308 AA: 34571 MW: 1D969187E63A2BAD CRC64:

Query Match Best Local Similarity 80.0%; Score 28; DB 1; Length 308;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;



OY 1 wrxxsfxx 8  
|||  
Db 103 WRQMSFWG 110



RESULT 6  
SERM_MYCLE STANDARD: PRT: 528 AA.  
AC O33116;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE D-3-PHOSPHOGLYCERATE DEHYDROGENASE (EC 1.1.1.95) (PCDH).  
GN SERA OR ML1692 OR MCB637.25.  
OS Mycobacterium leprae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
OX NCBI_TaxID=1769;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TN.  
RX MEDLINE-21128732; Pubmed-11234002;  
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,  
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
RA Murphy L., Oliver K., Quail M.A., Rajadream M.A., Rutherford K.M.,  
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
RA Barrell B.G.;  
RT "Massive gene decay in the leprosy bacillus."  
RL Nature 409:1007-1011(2001).  
CC -1- CATALYTIC ACTIVITY: 3-PHOSPHOGLYCERATE + NAD(+) =  
CC 3-PHOSPHOHYDROXYPYRUVATE + NADH.  
CC -1- PATHWAY: FIRST COMMITTED STEP IN THE "PHOSPHORYLATED" PATHWAY  
CC OF L-SERINE BIOSYNTHESIS.  
CC -1- SIMILARITY: BELONGS TO THE D-ISOMER SPECIFIC 2-HYDROXYACID  
CC DEHYDROGENASES FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).


```

EMBL: Z69263: CAB16440.1; -.
EMBL: AL563923; CAC30645.1; -.

```

DR HSSP: P01542; ICCN.
DR Leproma; ML1692; -.
DR InterPro: IPR002912; ACT.
DR InterPro: IPR002162; D_2_hydroxyacid_DH.
DR InterPro: IPR000205; NAD_binding.
DR Pfam: PF00389; 2-Hacid_DH; 1.
DR PROSITE: PS00065; D_2_HYDROXYACID_DH_1; 1.
DR PROSITE: PS00670; D_2_HYDROXYACID_DH_2; 1.
DR PROSITE: PS00671; D_2_HYDROXYACID_DH_3; 1.
DR Serine biosynthesis; Oxidoreductase; NAD; Complete proteome.
KW ACT_SITE 232 232 SUBSTRATE-BINDING (BY SIMILARITY).
FT ACT_SITE 261 261 BY SIMILARITY.
FT ACT_SITE 279 279 BY SIMILARITY.
SQ SEQUENCE 528 AA; 54469 MM; 1A6DC60F9FB71222 CRC64;

Query Match
Best Local Similarity 80.0%; Score 28; DB 1; Length 528;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 wrxsfxg 8
   1: 111
Db 129 WKRSSFSG 136

RESULT 7
SERB_MYCTU STANDARD; PRT; 528 AA.
AC 053243;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE D-3-PHOSPHOGLYCERATE DEHYDROGENASE (EC 1.1.1.95) (PGDH).
GN SERA OR RV2996C OR MT3074 OR MTV012.10.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
[1]
[2]
RA SEQUENCE FROM N.A.
RA STRAIN=H37RV;
RA MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne K., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellon S., Squares S., Squares R.,
RA Sultson J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RA complete genome sequence.";
RA Nature 393:537-544(1998).
[2]
RA SEQUENCE FROM N.A.
RA STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
RA laboratory strains.";
RA Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
[3]
RA -1- CATALYTIC ACTIVITY: 3-PHOSPHOGLYCERATE + NAD(+) =
RA -3- 3-PHOSPHOHYDROXYPYRUVATE + NADH.
RA -1- PATHWAY: FIRST COMMITTED STEP IN THE "PHOSPHORYLATED" PATHWAY
RA -1- OF L-SERINE BIOSYNTHESIS.
RA -1- SIMILARITY: BELONGS TO THE D-ISOMER SPECIFIC 2-HYDROXYACID
RA DEHYDROGENASES FAMILY.
RA -----
RA This Swiss-Prot entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL021287; CA116081.1; -.
DR EMBL: AE007127; AMK47403.1; -.
DR TIGR: MT3074; -.
DR Tuberculist; RV2996C; -.
DR InterPro: IPR002912; ACT.
DR InterPro: IPR002162; D_2_hydroxyacid_DH.
DR InterPro: IPR000205; NAD_binding.
DR Pfam: PF00389; 2-Hacid_DH; 1.
DR Pfam: PF01842; ACT; 1.
DR PROSITE: PS00065; D_2_HYDROXYACID_DH_1; 1.
DR PROSITE: PS00670; D_2_HYDROXYACID_DH_2; 1.
DR PROSITE: PS00671; D_2_HYDROXYACID_DH_3; 1.
DR Serine biosynthesis; Oxidoreductase; NAD; Complete proteome.
KW ACT_SITE 232 232 SUBSTRATE-BINDING (BY SIMILARITY).
FT ACT_SITE 261 261 BY SIMILARITY.
FT ACT_SITE 279 279 BY SIMILARITY.
SQ SEQUENCE 528 AA; 54554 MM; 3B5696AARD82A901 CRC64;

Query Match
Best Local Similarity 80.0%; Score 28; DB 1; Length 528;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 wrxsfxg 8
   1: 111
Db 129 WKRSSFSG 136

RESULT 8
EXSB_ECOLI STANDARD; PRT; 1180 AA.
ID EXSB_ECOLI
AC P08394;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE EXODEOXYRIBONUCLEASE V BETA CHAIN (EC 3.1.11.5) (EXODEOXYRIBONUCLEASE
DE V 135 KDA POLYPEPTIDE).
GN RECB OR ROBA OR B2820.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
[1]
RA SEQUENCE FROM N.A.
RA MEDLINE=87066729; PubMed=3537960;
RA Finch P.W., Storey A., Chapman K.E., Brown K., Hickson I.D.,
RA Emerson P.T.;
RA "Complete nucleotide sequence of the Escherichia coli recB gene.";
RA Nucleic Acids Res. 14:8573-8582(1986).
[2]
RA SEQUENCE FROM N.A.
RA STRAIN=V1000;
RA MEDLINE=20229837; PubMed=10766864;
RA Arnold D.A., Kowalczykowski S.C.;
RA "Facilitated loading of RecA protein is essential to recombination by
RA RecBCD enzyme.";
RA J. Biol. Chem. 275:12261-12265(2000).
[3]
RA SEQUENCE FROM N.A.
RA STRAIN=K12 / MG1655;
RA MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";

```

RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE OF 1-11 FROM N.A.
RA MEDLINE-87040734; PubMed-3534791;
RX Finch P.W., Wilson R.E., Brown K., Hickson I.D., Emerson P.T.;
RT "Complete nucleotide sequence of the Escherichia coli ptr gene
RL Nucleic Acids Res. 14:7695-7703(1986).
RN [5]
RP SEQUENCE OF 1093-1180 FROM N.A.
RX MEDLINE-87066730; PubMed-3537961;
RT Finch P.W., Storey A., Brown K., Hickson I.D., Emerson P.T.;
RL "Complete nucleotide sequence of recD, the structural gene for the
CC alpha subunit of Exonuclease V of Escherichia coli.";
RL Nucleic Acids Res. 14:8583-8594(1986).
CC -1- FUNCTION: REPAIRS FOR EFFICIENT DNA REPAIR: IT CATALYZES THE
CC UNWINDING OF DOUBLE-STRANDED DNA AND THE CLEAVAGE OF SINGLE-
CC STRANDED DNA AND IT STIMULATES LOCAL GENETIC RECOMBINATION.
CC ALL OF THESE ACTIVITIES REQUIRE CONCOMITANT HYDROLYSIS OF ATP.
CC -1- CATALYTIC ACTIVITY: EXONUCLEOTIC CLEAVAGE (IN THE PRESENCE OF
CC ATP) IN EITHER 5'-TO 3'-OR 3'-TO 5'-DIRECTION TO YIELD 5'-
CC PHOSPHOLICACIDOTIDES.
CC -1- SUBUNIT: CONSIST OF THREE SUBUNITS, RECD, RECD AND RECD.
CC -1- SIMILARITY: BELONGS TO THE UVRD SUBFAMILY OF HELICASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>
CC or send an email to license@isb-slb.ch).
CC -----
CC EMBL: X04581; CAA28250.1; -;
DR EMBL: AF179304; AAD56369.1; -;
DR EMBL: U29581; AAB40467.1; -;
DR EMBL: AB000365; AAC75859.1; -;
DR EMBL: X06227; CAA29577.1; -;
DR EMBL: X04582; CAA28252.1; -;
DR PIR: A25532; NCECX5.
DR HSSP: P56255; 1PUR.
DR Ecocore: BGI0824; recB.
DR InterPro: IPR000212; UVrd-helicase.
DR Pfam: PF00580; UVrd-helicase; 1.
KW Hydrolase; Nuclease; Exonuclease; Endonuclease; Helicase; ATP-binding;
KW DNA repair; Complete proteome.
FT NP_BIND 23 ATP.
FT SEQUENCE 1180 AA; 133958 MW; F9AC331808E8F281 CRC64;
OY 1 wrxxsfxy 8
DB 901 WRVTSYG 908
Query Match 80.0%; Score 28; DB 1; Length 1180;
Best Local Similarity 50.0%; Pred. No. 1,1e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OX NCBI_TaxID=9534;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-97242441; PubMed-9087440;
RX Katsuhira J., Inoue N., Horiguchi Y., Matsuda M., Sugimoto N.;
RT "Molecular cloning and functional characterization of the receptor for
RT Clostridium perfringens enterotoxin.";
RL J. Cell Biol. 136:1239-1247(1997).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>
CC or send an email to license@isb-slb.ch).
CC -----
CC EMBL: D88492; BAA22781.1; -;
DR InterPro: IPR001832; Claudin.
DR Pfam: PF00822; PMP22.Claudin.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN; 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 10 30
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 118 138 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
FT SEQUENCE 209 AA; 22029 MW; 474DB3099F95289E CRC64;
OY 1 wrxxsfxy 8
DB 30 WRVTAFIG 37
Query Match 77.1%; Score 27; DB 1; Length 209;
Best Local Similarity 50.0%; Pred. No. 36;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

RESULT 10
CLD4_HUMAN STANDARD; PRT; 209 AA.
ID CLD4_HUMAN
AC 014493;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE CLAUDIN-4 (CLOSTRIDIUM PERFRINGENS ENTEROTOXIN RECEPTOR) (CPE-
DE RECEPTOR) (CPE-R).
GN CLDN4 OR CPEPR1 OR CPER.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE-97476271; PubMed-9334247;
RA Katsuhira J., Sugiyama H., Inoue N., Horiguchi Y., Matsuda M.,
RT Sugimoto N.;
RT "Clostridium perfringens enterotoxin utilizes two structurally related
RT membrane proteins as functional receptors in vivo.";
RL J. Biol. Chem. 272:26652-26658(1997).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)

DR EMBL: AB000712; BAA22984.1; -.
DR MIM: 602909; -.
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin; 1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN; 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 118 138 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
SQ SEQUENCE 209 AA; 22077 MW; 0659A93AA5F0E4C5 CRC64;

Query Match 77.1%; Score 27; DB 1; Length 209;
Best Local Similarity 50.0%; Pred. No. 36;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 wrxsfxg 8
11:11
Db 30 WRVTAFIG 37

RESULT 11
CLD2_BRARE STANDARD; PRT; 209 AA.

AC O9YH90;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-LIKE PROTEIN ZF-A9.
OS Brachydanio rerio (Zebrafish) (Zebrafish).
OS Brachydanio rerio (Zebrafish) (Zebrafish).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;

RN [1]
RP SEQUENCE FROM N.A.
RA Keen T.J., Inglehearn C.F.;
RA Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)

DR EMBL: AJ011790; CAA09778.1; -.
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin; 1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN; 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 81 101 POTENTIAL.
FT TRANSMEM 114 134 POTENTIAL.
FT TRANSMEM 159 179 POTENTIAL.
SQ SEQUENCE 209 AA; 22091 MW; C413143811853D58 CRC64;

Query Match 77.1%; Score 27; DB 1; Length 209;

Best Local Similarity 50.0%; Pred. No. 36;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 wrxsfxg 8
11:11
Db 29 WRVTAFIG 36

RESULT 12
CLD4_MOUSE STANDARD; PRT; 210 AA.
ID CLD4_MOUSE
AC 035054;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-4 (CLOSTRIDIUM PERFRINGENS ENTEROTOXIN RECEPTOR) (CPE-
DE RECEPTOR) (CPE-R).
GN CLDN4 OR CPEPR1 OR CPER.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97476271; PubMed=9334247;
RA Katabira J., Sugiyama H., Inoue N., Horiguchi Y., Matsuda M.,
RA Sugimoto N.;
RT "Clostridium perfringens enterotoxin utilizes two structurally related
RT membrane proteins as functional receptors in vivo."
RL J. Biol. Chem. 272:26552-26558(1997).

RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99110921; PubMed=9892664;
RA Morita K., Furuse M., Fujimoto K., Tsukita S.;
RA "Claudin multigene family encoding four-transmembrane domain protein
RT components of tight junction strands."
RL Proc. Natl. Acad. Sci. U.S.A. 96:511-516(1999).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)

DR EMBL: AB000713; BAA22985.1; -.
DR EMBL: AF087822; AAD09757.1; -.
DR MGP: MG1:131314; CLD4.
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin; 1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN; 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
SQ SEQUENCE 210 AA; 22338 MW; 3B6D571E671D6564 CRC64;

Query Match 77.1%; Score 27; DB 1; Length 210;
Best Local Similarity 50.0%; Pred. No. 36;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 wrxsfxg 8
11:11
Db 30 WRVTAFIG 37

RESULT 13
CLD3_HUMAN STANDARD; PRT; 211 AA.
AC 095832;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CLAUDIN-1 (SENESCENCE-ASSOCIATED EPITHELIAL MEMBRANE PROTEIN).
OS CLDN1 OR CLD1 OR SEMP1.
OC Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE:99132301; PubMed=9931503;
RX MEDLINE:99132301; PubMed=9931503;
RA Swissbelm K.L., Machl A., Plantitzer S., Robertson R., Kubbies M.,
RA Hostler S.;
RT "SEMP1, a senescence-associated cDNA isolated from human mammary
RT epithelial cells, is a member of an epithelial membrane protein
RT superfamily.";
RL Gene 226:285-295(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Mitic L.M., Anderson J.M.;
RT "Human claudin-1 isolated from Caco-2 mRNA."
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE:20290992; PubMed=10828592;
RA Hallford S., Spencer P., Greenwood J., Winton H., Hunt D.M.,
RA Adamson P.;
RT "Assignment(1) of claudin-1 (CLDN1) to human chromosome 3q28-->q29
RT with somatic cell hybrids."
RL Cytogenet. Cell Genet. 88:217-217(2000).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF101051; AAD16433.1; -;
DR EMBL: AF115546; AAD22962.1; -;
DR EMBL: AF134160; AAF61393.1; -;
DR MIM: 603718; -;
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin.1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN.1.
DR TIGHT JUNCTION: Transmembrane.
KW TRANSMEM 8 28
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 116 136 POTENTIAL.
FT TRANSMEM 164 184 POTENTIAL.
FT CONFLICT 62 62 I -> V (IN REF. 2).
FT CONFLICT 135 135 V -> A (IN REF. 2).
SQ SEQUENCE 211 AA; 22744 MW; 07269000B6C214F0 CRC64;

Query Match 77.1%; Score 27; DB 1; Length 211;
Best Local Similarity 50.0%; Pred. No. 37;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 WRXSFSG 8
|| :| |

Db 30 WRISYAG 37
RESULT 14
CLD3_MOUSE STANDARD; PRT; 219 AA.
ID CLD3_MOUSE
AC 092069;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-3 (CLOSTRIDIUM PERRINGENS ENTEROTOXIN RECEPTOR 2) (CPE-
DE RECEPTOR 2) (CPE-R 2).
GN CLDN3 OR CPETR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:99097345; PubMed=9878248;
RX Paperna T., Peoples R., Wang Y.K., Kaplan P., Francke U.;
RT "Genes for the CPE receptor (CPETR1) and the human homolog of RVP1
RT (CPETR2) are localized within the Williams-Beuren syndrome deletion.";
RL Genomics 54:453-459(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE:99110921; PubMed=9892664;
RA Morita K., Furuse M., Fujimoto K., Tsukita S.;
RT "Claudin multigene family encoding four-transmembrane domain protein
RT components of tight junction strands."
RL Proc. Natl. Acad. Sci. U.S.A. 96:511-516(1999).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF095905; AAD14608.1; -;
DR EMBL: AF087821; AAD09756.1; -;
DR MGD: MGI:1329044; Cldn3.
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin.1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN.1.
DR TIGHT JUNCTION: Transmembrane.
KW TRANSMEM 9 29
FT TRANSMEM 81 101 POTENTIAL.
FT TRANSMEM 116 136 POTENTIAL.
FT TRANSMEM 160 180 POTENTIAL.
SQ SEQUENCE 219 AA; 23284 MW; 62F67810D9B9BD37 CRC64;

Query Match 77.1%; Score 27; DB 1; Length 219;
Best Local Similarity 50.0%; Pred. No. 38;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 WRXSFSG 8
|| :| |

Db 29 WRVSAFIG 36
RESULT 15
CLD3_RAT STANDARD; PRT; 219 AA.
ID CLD3_RAT
AC 063400;
DT 30-MAY-2000 (Rel. 39, Created)

```

DT 30-MAY-2000 (Rel. 39, last sequence update)
DT 30-MAY-2000 (Rel. 39, last annotation update)
DE CLAUDIN-3 (VENTRAL PROSTATE.1 PROTEIN) (RVP1).
GN CLDN3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92130987; PubMed=1723140;
RA Bilenl M.M., Miesfeld R.L.;
RT "Isolation and characterization of transcripts induced by androgen
RT withdrawal and apoptotic cell death in the rat ventral prostate.";
RL Mol. Endocrinol. 5:1381-1388(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Keen T.J., Inglehearn C.F.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M74067; AAA41760.1; -.
DR EMBL; AJ011656; CA09727.1; -.
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam; PF00822; PMP22_Claudin.
DR PRINTS; PR01077; CLAUDIN.
DR PROSITE; PS01346; CLAUDIN.1.
DR TIGHT junction; Transmembrane.
FT TRANSMEM 9 29 POTENTIAL.
FT TRANSMEM 81 101 POTENTIAL.
FT TRANSMEM 116 136 POTENTIAL.
FT TRANSMEM 162 182 POTENTIAL.
FT CONFLICT 4 4 G -> S (IN REF. 1).
FT CONFLICT 55 55 MISSING (IN REF. 1).
FT CONFLICT 217 219 DIV -> TTSRPGARPHHHNHYDSMPYTRPACSLASENT
PPSRRLQTPRSLARLEEDRDGVPFSPVAT (IN REF.
1).
SQ SEQUENCE 219 AA; 23314 MW; 820CC6BFC20D122D CRC64;

```

```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98110580; PubMed=9441748;
RA Peacock R.E., Keen T.J., Inglehearn C.F.;
RT "Analysis of a human gene homologous to rat ventral prostate.1
RT protein.";
RL Genomics 46:443-449(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97476271; PubMed=9334247;
RA Katahira J., Sugiyama H., Inoue N., Horiguchi Y., Matsuda M.,
RA Sugimoto N.;
RT "Clostridium perfringens enterotoxin utilizes two structurally related
RT membrane proteins as functional receptors in vivo.";
RL J. Biol. Chem. 272:26652-26658(1997)
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF007189; AAC78277.1; -.
DR EMBL; AB000714; BAA22986.1; -.
DR MIM; 602910; -.
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam; PF00822; PMP22_Claudin.1.
DR PRINTS; PR01077; CLAUDIN.
DR PROSITE; PS01346; CLAUDIN.1.
DR TIGHT junction; Transmembrane.
FT TRANSMEM 9 29 POTENTIAL.
FT TRANSMEM 81 101 POTENTIAL.
FT TRANSMEM 116 136 POTENTIAL.
FT TRANSMEM 160 180 POTENTIAL.
SQ SEQUENCE 220 AA; 23318 MW; IC826EFPF1563C56 CRC64;

```

```

Query Match 77.1%; Score 27; DB 1; Length 219;
Best Local Similarity 50.0%; Pred. No. 38;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 wxxsfxg 8
II : I I
Db 29 WRVSAFTG 36

```

```

Query Match 77.1%; Score 27; DB 1; Length 220;
Best Local Similarity 50.0%; Pred. No. 38;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 wxxsfxg 8
II : I I
Db 29 WRVSAFTG 36

```

HA Walter M., Fujlyama A., Taylor T.D., Watanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
 RA Ohki M., Takagi T., Sakaki Y., Tauden S., Blechschmidt K., Polley A.,
 RA Menzel U., Delabur J., Kumpf K., Lehmann R., Patterson D.,
 RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
 RA Rosenthal A., Kudon J., Shibuya K., Kawasaki K., Asakawa S.,
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
 RA Minoshima S., Shimizu N., Nordstiek G., Hornischer K., Brandt P.,
 RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
 RA Rammer J., Beck A., Klages S., Hennig S., Rlessolman L., Dagand E.,
 RA Wehmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
 RA Lehnach H., Reinhardt R., Yaspo M.-L.,
 RA "The DNA sequence of human chromosome 21." ;
 RA Nature 405:311-319(2000).
 CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AJ250712; CAB60616.1; -;
 CC EMBL: AP001707; BAA95566.1; -;
 CC InterPro: IPR001832; Claudin.
 DR InterPro: IPR000729; PMP22_Claudin.
 DR Pfam: PF00822; PMP22_Claudin.
 DR PRINTS: PS01077; CLAUDIN.
 DR PROSITE: PS01346; CLAUDIN; 1.
 DR TIGHT Junction: Transmembrane.
 KW TRANSMEM 8 28 POTENTIAL.
 FT TRANSMEM 82 102 POTENTIAL.
 FT TRANSMEM 125 145 POTENTIAL.
 FT TRANSMEM 165 185 POTENTIAL.
 SQ SEQUENCE 224 AA: 24603 MW: 1833ED3178B7F63A CRC64;

Query Match 77.1%; Score 27; DB 1; Length 224;
 Best Local Similarity 50.0%; Pred. No. 39;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfxxg 8
 || | |
 DB 30 WRVSAPVG 37

RESULT 18
 SPS2_HUMAN STANDARD; PRT; 448 AA.
 AC 099611;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE SELENIDE WATER DIFORMSE 2 (EC 2.7.9.3) (SELENOPHOSPHATE SYNTHETASE 2)
 DE (SELENIUM DONOR PROTEIN 2).
 GN SPS2.
 OS Homo sapiens (Human).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE:96017645; PubMed: 7588067;
 RA Guimaraes M.J., Bazan J.F., Zlotnik A., Wiles M.V., Grimaldi J.C.,
 RA Lee F., McLanahan T.;
 RT "A new approach to the study of haematopoietic development in the
 RT yolk sac and embryoid bodies";
 RL Development 121:3335-3346(1995).
 RN [2]

RP SEQUENCE FROM N.A.
 RA MEDLINE:97140286; PubMed:8986768;
 RA Guimaraes M.J., Peterson D., Vicari A., Cocks B.G., Copeland N.G.,
 RA Gilbert D.J., Jenkins N.A., Ferlic D.A., Kostelet R., Bazan J.F.,
 RA Zlotnik A.;
 RT "Identification of a novel seld homolog from eukaryotes, bacteria,
 RT and archaea: is there an autoregulatory mechanism in selenocysteine
 RT metabolism?";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:15086-15091(1996).
 CC -1- FUNCTION: SYNTHESIZES SELENOPHOSPHATE FROM SELENIDE AND ATP.
 CC -1- CATALYTIC ACTIVITY: ATP + SELENIDE + H(2)O = AMP + SELENOPHOSPHATE
 CC + PHOSPHATE.
 CC -1- COFACTOR: SELENOCYSTEINE. THE ACTIVE-SITE SELENOCYSTEINE IS
 CC ENCODED BY THE OPAL CODON, UGA.
 CC -1- SIMILARITY: BELONGS TO THE SELENOPHOSPHATE SYNTHETASE 1 FAMILY.
 CC CLASS I SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U43286; AAC50958.1; -;
 CC MIM: 606218; -;
 DR InterPro: IPR00728; ATRS-related.
 DR Pfam: PF00586; ATRS; 1.
 KW Transferase: Selenium; Selenocysteine; ATP-binding.
 FT ACT_SITE 60 60
 FT SE_CYS 60 60
 FT SITE 63 63
 FT NP_BIND 319 325 ATP (POTENTIAL)
 SQ SEQUENCE 448 AA: 47258 MW: 343A58CD9F842B99 CRC64;

Query Match 77.1%; Score 27; DB 1; Length 448;
 Best Local Similarity 50.0%; Pred. No. 73;
 Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxxsfxxg 8
 || | |
 DB 49 WRLTGFSG 56

RESULT 19
 ADRO_DROME STANDARD; PRT; 466 AA.
 ID ADRO_DROME
 AC 09V379;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NADPH:ADRENODOXIN OXIDOREDUCTASE, MITOCHONDRIAL PRECURSOR
 DE (EC 1.18.1.2) (ADRENODOXIN REDUCTASE) (AR) (FERREDOXIN-NADP(+)
 DE REDUCTASE).
 GN DARE OR CG12390.
 OS Drosophila melanogaster (Fruit fly).
 OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
 RC STRAIN=W1118; TISSUE=Testis, and Head;
 RA MEDLINE:99429818; PubMed:10498693;
 RA Freeman M.R., Dobritsa A., Gaines P., Segraves W.A., Carlson J.R.;
 RT "The dare gene: steroid hormone production, olfactory behavior, and
 RT neural degeneration in Drosophila.";
 RL Development 126:4591-4602(1999).
 RN [2]

CC STRAIN=BERKELEY; PubMed-10731132;
RX MEDLINE=20196006;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agayau A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernier B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstedt P., Brothier P.,
RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dutton K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitani M., Kalush F., Karp G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA Lasok P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Pui Y., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Splier E., Spirding A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Weissman D.A., Weisskock G.M., Weissbach J.,
RA Williams S.M., Woodard J., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
CC -I- FUNCTION: REQUIRED FOR SYNTHESIS OF STEROID HORMONES. FOR
CC OLFACTORY SENSORY BEHAVIOR AND COMPLETION OF THE SECOND LARVAL
CC MOLT (A STEROID MEDIATED DEVELOPMENTAL TRANSITION) AND
CC POUARIATION.
CC
CC -I- CATALYTIC ACTIVITY: REDUCED ADRENODOXIN + NAD(P)+ = OXIDIZED
CC ADRENODOXIN + NADPH.
CC
CC -I- COFACTOR: FAD.
CC
CC -I- PATHWAY: CHOLESTEROL SIDE-CHAIN-CLEAVAGE SYSTEM.
CC
CC -I- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC
CC -I- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN PROTHORACIC GLAND
CC OF THE LARVAL RING GLAND AND NURSE CELLS OF THE ADULT OVARY. LOW
CC EXPRESSION IS ALL ADULT TISSUES EXAMINED.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
CC or send an email to license@1sb-sib.ch).
CC
CC
CC EMBL, AF168685; AAD50819.1; -
DR EMBL, AE003826; AAF58678.1; -
DR HSSP, P08165; 1C0C.
DR FLYBase: FBgn0015582; dare.
DR InterPro: IPR000759; Adrxndx_redctse.
DR InterPro: IPR00103; pyridine_redox_2.
DR PRINTS: PR00419; ADXRDTASE.
DR PRINTS: PR00469; PNDRTASEII.
KM Electron transport; Oxidoreductase; Flavoprotein; NAD; FAD;
KM Mitochondrion; Transit peptide.
TM Transist 1 ?
CT CHAIN ? 466 MITOCHONDRION (POTENTIAL).
NADPH:ADRENODOXIN OXIDOREDUCTASE.

SQL	SEQUENCE	466 AA;	51352 MM;	6AD7C1CF0F1CAFD6 CRC64;
	Query Match		77.1%;	Score 27; DB 1; Length 466;
	Best Local Similarity		50.0%;	Pred. No. 75;
	Matches	4; Conservative	0; Mismatches	4; Indels
			0; Gaps	0;
QY	1 wxxxxfxxg 8			
Db	246 WRTEDFSG 253			
RESULT	20			
YVD3_CAEEL				
ID	YVD3_CAEEL	STANDARD;	PRT;	551 AA.
AC	P55114;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DE	20-AUG-2001 (Rel. 40, Last annotation update)			
GN	K04E7.3.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;			
CC	Rhabditidae; Pelodierinae; Caenorhabditis.			
NCBI_Taxid=6239;				
SEQUENCE FROM N.A.				
RC	STRAIN-BRISTOL N2;			
RA	Nhan M.;			
RL	Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.			
CC	-I- SIMILARITY: CONTAINS 1 CUB DOMAIN.			
CC	-I- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.			
CC	-I- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.			
CC	-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12A (ZINC METALLOPROTEINASE); ALSO KNOWN AS THE ASTACIN SUBFAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@lsb-sib.ch).			
CC	-----			
DR	EMBL; U39666; AAA80412.1; -.			
DR	HSSP; P00740; 11XA.			
DR	WormPepe; K04E7.3; CE02798.			
DR	InterPro; IPR001506; Astacin.			
DR	InterPro; IPR000859; CUB.			
DR	InterPro; IPR000561; EGF-like.			
DR	InterPro; IPR000884; TSP1.			
DR	InterPro; IPR000130; Zn.MTpeptidse.			
DR	Pfam; PF01400; Astacin; 1.			
DR	Pfam; PF00431; CUB; 1.			
DR	SMART; SM00042; CUB; 1.			
DR	SMART; SM00209; TSP1; 1.			
DR	SMART; SM00235; Zmc; 1.			
DR	PROSITE; PS01180; CUB; FALSE_NEG.			
DR	PROSITE; PS00022; EGF_1; 1.			
DR	PROSITE; PS01186; EGF_2; FALSE_NEG.			
DR	PROSITE; PS50092; TSP1; 1.			
DR	PROSITE; PS00142; ZINC_PROTEASE; 1.			
KM	Hypothetical protein; Hydrolase; Metalloprotease; Zmc;			
KM	Signal; EGF-like domain; Glycoprotein.			
FT	SIGNAL	1	?	POTENTIAL.
FT	CHAIN	?	551	HYPOTHETICAL ZINC METALLOPROTEINASE
FT				K04E7.3.
FT	DOMAIN	18	28	POLY-PRO.
FT	DOMAIN	318	358	EGF-LIKE.
FT	DOMAIN	359	484	CUB.
FT	DOMAIN	485	535	TSP TYPE-1.
FT	METAL	219	219	ZINC (CATALYTIC) (BY SIMILARITY).
FT	ACT_SITE	220	220	BY SIMILARITY.

FT METAL 223 223 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 229 229 ZINC (CATALYTIC) (BY SIMILARITY).
FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 378 378 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 551 AA; 61673 MW; 7CFC9A16B56C87E CRC64;

Query Match 77.1%; Score 27; DB 1; Length 551;
Best Local Similarity 50.0%; Pred. No. 88;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wtxsfxg 8
DB 376 WRNISYSG 383

Search completed: January 14, 2002, 07:40:40
Job time: 506 sec

Gencore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:39:47 ; Search time 112.89 Seconds
(without alignments)
10.366 Million cell updates/sec

Title: 09-185908-1g
Perfect score: 35
Sequence: 1 wrxsfxg 8

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 100%
Listing first 50 summaries

Database :
1: SP-archaea:*
2: SP-bacteria:*
3: SP-fungi:*
4: SP-human:*
5: SP-invertebrate:*
6: SP-mammal:*
7: SP-mhc:*
8: SP-organelle:*
9: SP-phage:*
10: SP-plant:*
11: SP-rodent:*
12: SP-virus:*
13: SP-vertebrate:*
14: SP-unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	31	88.6	206 2	Q9RBD0
2	30	85.7	1083 4	Q14868
3	30	85.7	1083 11	Q63744
4	30	85.7	1091 4	O43199
5	30	85.7	1092 11	Q9R029
6	30	85.7	1554 4	Q9C0E0
7	29	82.9	149 5	Q9NM13
8	29	82.9	777 3	Q74636
9	29	82.9	1715 10	Q9S1S9
10	28	80.0	65 2	Q9RDQ2
11	28	80.0	65 2	Q9RD18
12	28	80.0	65 2	Q9RD17
13	28	80.0	197 11	Q9JH60
14	28	80.0	243 2	Q9ABD0
15	28	80.0	255 8	Q9A394
16	28	80.0	266 2	Q9EWM0
17	28	80.0	1029 2	Q9A9L0
18	28	80.0	1208 2	Q9KPP6
19	28	80.0	1778 5	Q9NE65

20	27	77.1	59 4	Q9BU02	Q9BU02 homo sapien
21	27	77.1	115 2	Q9CND1	Q9CND1 pasteurilla
22	27	77.1	159 9	Q38144	Q38144 bacterioph
23	27	77.1	212 2	Q49799	Q49799 mycobacteri
24	27	77.1	214 13	Q9DE12	Q9DE12 xenopus lae
25	27	77.1	214 13	Q9SR2	Q9SR2 gallus gall
26	27	77.1	278 1	Q58458	Q58458 pyrococcus
27	27	77.1	294 2	Q9S377	Q9S377 mycobacteri
28	27	77.1	323 5	Q9N7R6	Q9N7R6 leishmania
29	27	77.1	329 10	Q9SU8	Q9SU8 arabidopsis
30	27	77.1	330 2	Q9S2V4	Q9S2V4 streptomyce
31	27	77.1	344 5	Q18417	Q18417 caenorhabd
32	27	77.1	370 2	Q9HY01	Q9HY01 pseudomonas
33	27	77.1	372 2	Q9RD05	Q9RD05 pseudomonas
34	27	77.1	460 10	Q9S7R7	Q9S7R7 arabidopsis
35	27	77.1	484 10	Q9X1Y1	Q9X1Y1 oryza sativ
36	27	77.1	530 2	Q9AC16	Q9AC16 caulobacter
37	27	77.1	619 5	Q9VA04	Q9VA04 drosophila
38	27	77.1	735 5	Q9W0L8	Q9W0L8 drosophila
39	27	77.1	895 10	Q9FGN6	Q9FGN6 arabidopsis
40	27	77.1	998 2	Q9HU13	Q9HU13 pseudomonas
41	26	74.3	85 2	P95006	P95006 mycobacteri
42	26	74.3	122 2	Q9PHJ9	Q9PHJ9 xylella fas
43	26	74.3	141 1	Q9XCN6	Q9XCN6 aeropyrum p
44	26	74.3	154 2	O50394	O50394 mycobacteri
45	26	74.3	182 2	P97158	P97158 escherichia
46	26	74.3	229 4	Q9H5X9	Q9H5X9 homo sapien
47	26	74.3	231 10	Q9ZWS9	Q9ZWS9 arabidopsis
48	26	74.3	239 12	Q83736	Q83736 beet wester
49	26	74.3	239 12	Q9IMU2	Q9IMU2 beet wester
50	26	74.3	239 12	Q9IMU1	Q9IMU1 beet wester

ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	206 AA.
Q9RBD0	Q9RBD0			
AC	Q9RBD0			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	3-DEMTHTYLBUIQUNONE-9-3-METHYLTFRANSFERASE, PUTATIVE.			
GN	DR2562.			
OS	Deinococcus radiodurans.			
OC	Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.			
OX	NCBI_TaxID=1299;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=R1;			
RX	MEDLINE=20036896; PubMed=10567266;			
RA	White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,			
RA	Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,			
RA	Vamathevan J.J., Lam P., McDonald L., Uteerback T., Zaleski C.,			
RA	Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,			
RA	Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,			
RT	Fraser C.M.;			
RT	"Genome sequence of the radioresistant bacterium Deinococcus			
RT	radiodurans R1.";			
RL	Science 286:1571-1577(1999).			
DR	EMBL: AE002085; AAF12100.1; -.			
DR	TIGR: DR2562; -.			
DR	InterPro: IPR001601; Meth-transf.			
DR	InterPro: IPR000051; SAM bind.			
KW	Transferase; Methyltransferase; Ubiquitome; Complete proteome.			
SQ	SEQUENCE 206 AA; 22186 MW; 6F63E1369E12D870 CRC64;			

Query Match 88.6%; Score 31; DB 2; Length 206;
Best Local Similarity 62.5%; Pred. No. 22;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

OY      1 wrxxsfxx 8
      |||||
Db      111 WRTGSFHC 118

RESULT 2
ID 014868 PRELIMINARY; PRT; 1083 AA.
AC 014868.
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HP PROTEIN.
GN HP
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Wei M.-H., Pack S., Ivanov S., Lerman M.I.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF026219; AAB81637.1; -.
DR HSSP; Q07960; IRGP.
DR InterPro; IPR001005; Myb_DNA_bind.
DR InterPro; IPR00198; RhogAP.
DR InterPro; IPR001660; SAM.
DR InterPro; IPR002913; START.
DR Pfam; PF00620; RhogAP; 1.
DR Pfam; PF01852; START; 1.
DR SMART; SM00324; RhogAP; 1.
DR SMART; SM00454; SAM; 1.
DR SMART; SM00234; START; 1.
DR PROSITE; PS00037; MYB_1; UNKNOWN_1.
SQ SEQUENCE 1083 AA; 121897 MW; 6C1601F312749AA8 CRC64;

Query Match      85.7%; Score 30; DB 4; Length 1083;
Best Local Similarity 62.5%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY      1 wrxxsfxx 8
      |||||
Db      393 WRTGSFHC 400

RESULT 3
ID 063744 PRELIMINARY; PRT; 1083 AA.
AC 063744.
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE RHOGAP PROTEIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN;
RX MEDLINE=95137008; PubMed=7835339;
RA Homma Y., Emori Y.;
RT "A dual functional signal mediator showing RhogAP and phospholipase C-
RT delta stimulating activities.";
RL EMBO J. 14:286-291(1995).
DE DELETED IN LIVER CANCER 1.
DE HSSP; Q07960; IRGP.
DR InterPro; IPR001005; Myb_DNA_bind.
DR InterPro; IPR00198; RhogAP.
DR InterPro; IPR001660; SAM.

```

```

DR InterPro; IPR002913; START.
DR Pfam; PF00620; RhogAP; 1.
DR Pfam; PF01852; START; 1.
DR SMART; SM00324; RhogAP; 1.
DR SMART; SM00454; SAM; 1.
DR SMART; SM00234; START; 1.
DR PROSITE; PS00037; MYB_1; UNKNOWN_1.
SQ SEQUENCE 1083 AA; 122465 MW; 076D0E77ACD9D012 CRC64;

Query Match      85.7%; Score 30; DB 11; Length 1083;
Best Local Similarity 62.5%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY      1 wrxxsfxx 8
      |||||
Db      393 WRTGSFHC 400

RESULT 4
ID 043199 PRELIMINARY; PRT; 1091 AA.
AC 043199.
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE DELETED IN LIVER CANCER 1.
GN DLC-1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Yuan B.Z., Miller M.J., Keck C.L., Zimonjic D.B., Thorgerirsson S.S.,
RA Popescu N.C.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF035119; AAB87700.1; -.
DR HSSP; Q07960; IRGP.
DR InterPro; IPR001005; Myb_DNA_bind.
DR InterPro; IPR00198; RhogAP.
DR InterPro; IPR001660; SAM.
DR InterPro; IPR002913; START.
DR Pfam; PF00620; RhogAP; 1.
DR Pfam; PF01852; START; 1.
DR SMART; SM00324; RhogAP; 1.
DR SMART; SM00454; SAM; 1.
DR SMART; SM00234; START; 1.
DR PROSITE; PS00037; MYB_1; UNKNOWN_1.
SQ SEQUENCE 1091 AA; 122816 MW; 51712DE7ECD0F52A CRC64;

Query Match      85.7%; Score 30; DB 4; Length 1091;
Best Local Similarity 62.5%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY      1 wrxxsfxx 8
      |||||
Db      401 WRTGSFHC 408

RESULT 5
ID 09R029 PRELIMINARY; PRT; 1092 AA.
AC 09R029.
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE DELETED IN LIVER CANCER 1.
GN ARHGAP7 OR DLC-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```


OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Yuan B.Z., Yang Y., Keck C.L., Zimonjic D.B., Thorgeirsson S.S.,
RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF178078; AAD51760.1; -.
DR HSSP: Q07960; IRCP.
DR MGD: MGI:1354949; Arhgap7.
DR InterPro: IPR001005; Myb_DNA_bind.
DR InterPro: IPR00198; RhogAP.
DR InterPro: IPR001660; SAM.
DR InterPro: IPR002913; START.
DR Pfam: PF00620; RhogAP; 1.
DR Pfam: PF01852; START; 1.
DR SMART: SM00324; RhogAP; 1.
DR SMART: SM00454; SAM; 1.
DR SMART: SM00234; START; 1.
DR PROSITE: PS00037; MYB_1; UNKNOWN_1.
SQ SEQUENCE 1092 AA; 123390 MW; 2355B11D70C85046 CRC64;

Query Match 85.7%; Score 30; DB 11; Length 1092;
Best Local Similarity 62.5%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxsfxg 8
ID 402 WRTGSFHG 409

RESULT 6
ID 09C0E0 PRELIMINARY; PRT; 1554 AA.
AC 09C0E0;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE KIAA1723 PROTEIN (FRAGMENT).
CN KIAA1723.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21082932; PubMed=11214970;
RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
DR DNA Res. 7:347-355(2000).
DR EMBL: AB051510; BAB21814.1; -.
FT NON_TER 1
SQ SEQUENCE 1554 AA; 173549 MW; 76FD31F2139F2E12 CRC64;

Query Match 85.7%; Score 30; DB 4; Length 1554;
Best Local Similarity 62.5%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxsfxg 8
ID 864 WRTGSFHG 871

RESULT 7
ID 09NM13 PRELIMINARY; PRT; 149 AA.
AC 09NM13;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)

DE PROBABLE HYPOTHETICAL 15.1 KDA PROTEIN IN HUBB-COF INTERGENIC REGION
DE (FRAGMENT).
GN LM26.109.
OS Leishmania major.
OC Leishmania; Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIDELIN;
RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL160493; CAB97727.1; -.
FT NON_TER 149
SQ SEQUENCE 149 AA; 16592 MW; 62B7B040A69F2721 CRC64;

Query Match 82.9%; Score 29; DB 5; Length 149;
Best Local Similarity 62.5%; Pred. No. 44;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxsfxg 8
ID 28 WRQLSFLG 35

RESULT 8
ID 074636 PRELIMINARY; PRT; 777 AA.
AC 074636;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE TRANSPOSASE-LIKE PROTEIN.
OS Fusarium oxysporum.
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Fusarium.
OX NCBI_TaxID=5507;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF305118; TRANSPOSON=AC-TYPE TRANSPOSON, TFOL;
RX MEDLINE=9834103; PubMed=9671028;
RA Okuda M., Ikeda K., Nishiki F., Tsuge T.;
RT "TfOl: an Ac-like transposon from the plant pathogenic fungus Fusarium
RT oxysporum.";
RT Mol. Gen. Genet. 258:599-607(1998).
DR EMBL: AB008746; BAA32244.1; -.
DR InterPro: IPR003656; BED_finger.
SQ SEQUENCE 777 AA; 88602 MW; 5A067280645B836E CRC64;

Query Match 82.9%; Score 29; DB 3; Length 777;
Best Local Similarity 62.5%; Pred. No. 2.2e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxsfxg 8
ID 402 WRKOSFIG 409

RESULT 9
ID 09SIS9 PRELIMINARY; PRT; 1715 AA.
AC 09SIS9;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE PUTATIVE NON-LTR RETROELEMENT REVERSE TRANSCRIPTASE.
GN AT2G01840.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Kounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buehl C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Crozin L.A., Shen M., Vanaken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhagen G.P., Preuss D., Nieman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RL "Sequence and analysis of chromosome II of Arabidopsis thaliana.";
RL Nature 402:761-768(1999).
CC -I SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
TRANSCRIPTASE).
CC EMBL: AC007069; A021778.1; -
DR InterPro: IPR002203; Inein.
DR InterPro: IPR000477; RVse.
DR InterPro: IPR000130; Zn_MTPptase.
DR Pfam: PF00078; rvt.1.
DR PROSITE: PS00881; PROTEIN_SPLICING; UNKNOWN_1.
DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR RNA-directed DNA polymerase.
KW SEQUENCE 1715 AA; 198103 MW; CC88194C805F3830 CRC64;

OY 1 wrxxsfxxg 8
|||
Db 1450 WRSANFSG 1457

RESULT 10
O9RD02 PRELIMINARY; PRT; 65 AA.
AC O9RD02;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE HYPOTHETICAL 6.9 KDA PROTEIN.
GN SC4A7.10.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL: AL133423; CAB62714.1; -;
KW Hypothetical protein.
SQ SEQUENCE 65 AA; 6944 MW; F283FA15A0650DCE CRC64;

Query Match 80.0%; Score 28; DB 2; Length 65;
Best Local Similarity 50.0%; Pred. No. 32;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfxxg 8
|||
Db 12 WRRSYSG 19

RESULT 11
O9RD18 PRELIMINARY; PRT; 65 AA.
AC O9RD18;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE HYPOTHETICAL 7.0 KDA PROTEIN.
GN SCC57A.09C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL: AL136519; CAB62717.1; -;
KW Hypothetical protein.
SQ SEQUENCE 65 AA; 6959 MW; 1F74C265B9572610 CRC64;

Query Match 80.0%; Score 28; DB 2; Length 65;
Best Local Similarity 50.0%; Pred. No. 32;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfxxg 8
|||
Db 7 WRRSYSG 14

RESULT 12
O9RD17 PRELIMINARY; PRT; 65 AA.
AC O9RD17;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE HYPOTHETICAL 7.1 KDA PROTEIN.
GN SCC57A.10C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RC MEDLINE:97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RW EMBL: AL136519; CAB66278.1; -.
KW Hypothetical protein.
SQ SEQUENCE 65 AA; 7127 MW; 4C9A4476C44B727A CRC64;

Query Match 80.0%; Score 28; DB 2; Length 65;
Best Local Similarity 50.0%; Pred. No. 32;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfxxg 8
| | | |
Db 7 WKXSTSG 14

RESULT 13
O9JHGO PRELIMINARY; PRT; 197 AA.
AC O9JHGO;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CBLN3.
GN CBLN3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Pang Z., Morgan J.I.;
RT "Cloning and characterization of a novel precerebellin-related gene.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF218380; AAF32315.1; -.
DR EMBL: AF218379; AAF32314.1; -.
DR MGD; MGI:1889286; Cbln3.
DR InterPro; IPR001073; Clq.
DR Pfam; PF00386; Clq; 1.
DR PRINTS; PRO0007; COMPLEMENTC1Q.
DR SMART; SM00110; C1Q; 1.
DR PROSITE; PS01113; C1Q; 1.
SQ SEQUENCE 197 AA; 21077 MW; DBA8925C9BB1B77 CRC64;

Query Match 80.0%; Score 28; DB 11; Length 197;
Best Local Similarity 50.0%; Pred. No. 95;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfxxg 8
| | | |
Db 184 WKXSFSG 191

RESULT 14
O9ADD0 PRELIMINARY; PRT; 243 AA.
ID O9ADD0;
AC O9ADD0;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
SQ SEQUENCE 65 AA; 7127 MW; 4C9A4476C44B727A CRC64;
```

```

DE HYPOTHETICAL 27.3 KDA PROTEIN.
GN SCBAC5H2.06C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RC MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL: AL589707; CAC33901.1; -.
KW Hypothetical protein.
SQ SEQUENCE 243 AA; 27350 MW; 819C5DF7366DBE70 CRC64;

Query Match 80.0%; Score 28; DB 2; Length 243;
Best Local Similarity 50.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfxxg 8
| | | |
Db 47 WKASFLG 54

RESULT 15
O9B394 PRELIMINARY; PRT; 255 AA.
ID O9B394;
AC O9B394;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CYTOCHROME B (FRAGMENT).
GN CYTB.
OS Blackburnia palmeae.
OC Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Aedeptera; Carabidae;
OC Carabidae; Carabidae; Coleoptera; Aedeptera; Carabidae;
OX NCBI_TaxID=155379;
RN [1]
RP SEQUENCE FROM N.A.
RA Cryan J.R., Liebherr J.K., Fetzner J.W. Jr., Whitting M.F.;
RT "Evaluation of relationships within the endemic Hawaiian Platynini
RT (Coleoptera: Carabidae) based on molecular and morphological
evidence.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF338091; AAK28197.1; -.
KW Mitochondrion.
FT NON_TER 1 1
FT NON_TER 255 255
SQ SEQUENCE 255 AA; 29021 MW; 0BBAA66465004D11 CRC64;

Query Match 80.0%; Score 28; DB 8; Length 255;
Best Local Similarity 62.5%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfxxg 8
```

Db 85 WROMSFWG 92

RESULT 16

09EMW0 ID 09EMW0 PRELIMINARY; PRT; 266 AA.
AC 09EMW0;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE PUTATIVE ENOYL-COA HYDRATASE.
GN 2SCK31.1IC.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RL "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL MOL. Microbiol. 21:77-96(1996).
DR EMBL; AL512667; CAC21620.1; -.
DR InterPro: IPR001753; ENOYL_COA_hydrase.
DR Pfam: PF00378; ECH; 1.
DR PROSITE: PS00166; ENOYL_COA_HYDRATASE; 1.
SQ SEQUENCE 266 AA; 28121 MW; C07P9346B82E0451 CRC64;

Query Match 80.0%; Score 28; DB 2; Length 266;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;

Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxsfxg 8
Db 239 WRSVAFSG 246

RESULT 17

09A9L0 ID 09A9L0 PRELIMINARY; PRT; 1029 AA.
AC 09A9L0;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE TONB-DEPENDENT RECEPTOR.
GN CC0970.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phade N.D., Ely B.,
RA Deboy R.T., Dodson R.J., Durkin A.S., Gwin M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,

RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RL "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE05774; AAK22954.1; -.

DR TIGR; CC0970; -.
KW Receptor; Complete proteome.
SQ SEQUENCE 1029 AA; 111579 MW; 9EB952CF35705847 CRC64;

Query Match 80.0%; Score 28; DB 2; Length 1029;
Best Local Similarity 50.0%; Pred. No. 4.9e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxsfxg 8
Db 804 WRSNFDG 811

RESULT 18

09KPP6 ID 09KPP6 PRELIMINARY; PRT; 1208 AA.
AC 09KPP6;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE EXODEOXYRIBONUCLEASE V, 135 KDA SUBUNIT.
GN VC2320.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Hodelson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettein H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RL "DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae.";
RL Nature 406:477-483(2000).
DR EMBL; AE004302; AAF95464.1; -.
DR TIGR; VC2320; -.
DR InterPro: IPR000212; UVRD-helicase.
DR Pfam: PF00580; UVRD-helicase; 2.
KW Complete proteome.
SQ SEQUENCE 1208 AA; 135860 MW; F3C3EF891435C18D CRC64;

Query Match 80.0%; Score 28; DB 2; Length 1208;
Best Local Similarity 50.0%; Pred. No. 5.7e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxsfxg 8
Db 906 WRVTSYG 913

RESULT 19

09NE65 ID 09NE65 PRELIMINARY; PRT; 1778 AA.
AC 09NE65;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, last annotation update)
DE HYPOTHETICAL 183.5 KDA PROTEIN.
GN L7758.02.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.

OX NCBI_TaxID:5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Rieger M., Fuchs M., Gabel C., Mueller-Auer S., Schaefer M.,
RA Ivens A.C., Quail M., Rajandream M.A., Barrell B.G.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RX MEDLINE:98146435; PubMed:9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
DR EMBL: AL352980; CAB88218.1; -;
KW Hypothetical Protein.
SQ SEQUENCE 1778 AA; 183516 MW; AAC6C69BDF6B30D CRC64;

Query Match 80.0%; Score 28; DB 5; Length 1778;
Best Local Similarity 62.5%; Pred. No. 8.3e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxsfxg 8
DB 1309 WRWSPFG 1316

RESULT 20
ID 09BU02 PRELIMINARY; PRF; 59 AA.
AC 09BU02;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE SELENOPHOSPHATE SYNTHETASE 2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RHABDOMYOSARCOMA;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC002381; AAH02381.1; -;
SQ SEQUENCE 59 AA; 5897 MW; 8DFC9070558436D5 CRC64;

Query Match 77.1%; Score 27; DB 4; Length 59;
Best Local Similarity 50.0%; Pred. No. 48;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wxxsfxg 8
DB 49 WRLTGFSG 56

Search completed: January 14, 2002, 07:39:49
Job time: 960 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:36:01 ; Search time 103.51 Seconds
(without alignments)
5.725 Million cell updates/sec

Title: 09-185908-1g
Perfect score: 35
Sequence: 1 wrxxsfxy 8

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database :
1: /SID52/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	85.7	1091	20	AAV30337
2	29	82.9	41	22	AA871661
3	29	82.9	120	21	AA853841
4	29	82.9	417	21	AA807679
5	28	80.0	8	21	AA806521
6	28	80.0	8	21	AA806574
7	28	80.0	10	21	AA806530
8	28	80.0	10	21	AA806539
9	28	80.0	10	21	AA806548
10	28	80.0	10	21	AA806557
11	28	80.0	10	21	AA806566

12	28	80.0	59	18	AAW20137	H. pylori cytoplasm
13	28	80.0	205	22	AAV99420	Human PRO1486 (UNQ)
14	28	80.0	205	22	AA866169	Protein of the Inv
15	28	80.0	206	20	AAW88747	Secreted protein e
16	28	80.0	230	21	AAV51676	Murine clodin 2 pr
17	27	77.1	8	21	AA806426	Claudin-1 cell adh
18	27	77.1	8	21	AA806512	Claudin-1 cell adh
19	27	77.1	8	21	AA806583	Claudin-3 cell adh
20	27	77.1	8	21	AA806636	Claudin-3 cell adh
21	27	77.1	8	21	AA806645	Claudin-4 cell adh
22	27	77.1	9	21	AA806698	Claudin-1 cell adh
23	27	77.1	10	21	AA806427	Claudin-1 cell adh
24	27	77.1	10	21	AA806485	Claudin-1 cell adh
25	27	77.1	10	21	AA806491	Claudin-1 cell adh
26	27	77.1	10	21	AA806497	Claudin-1 cell adh
27	27	77.1	10	21	AA806503	Claudin-1 cell adh
28	27	77.1	10	21	AA806509	Claudin-1 cell adh
29	27	77.1	10	21	AA806592	Claudin-3 cell adh
30	27	77.1	10	21	AA806601	Claudin-3 cell adh
31	27	77.1	10	21	AA806610	Claudin-3 cell adh
32	27	77.1	10	21	AA806619	Claudin-3 cell adh
33	27	77.1	10	21	AA806628	Claudin-3 cell adh
34	27	77.1	10	21	AA806654	Claudin-4 cell adh
35	27	77.1	10	21	AA806663	Claudin-4 cell adh
36	27	77.1	10	21	AA806672	Claudin-4 cell adh
37	27	77.1	10	21	AA806681	Claudin-4 cell adh
38	27	77.1	10	21	AA806690	Claudin-4 cell adh
39	27	77.1	71	20	AAV38421	Human secreted pro
40	27	77.1	78	21	AA843420	Human cancer assoc
41	27	77.1	202	20	AAW86629	Secreted protein e
42	27	77.1	208	21	AA852100	Gene 48 human secr
43	27	77.1	209	21	AA843133	Human ORFX ORF2897
44	27	77.1	210	22	AAW866107	Kidney injury asso
45	27	77.1	210	22	AAW866107	Human colon cancer
46	27	77.1	211	20	AAV41726	Human PRO944 prote
47	27	77.1	211	20	AAV04143	Human Tango-73 pro
48	27	77.1	211	20	AAW96653	Human senescence f
49	27	77.1	211	21	AA844282	Human PRO944 (UNQ)
50	27	77.1	211	21	AAV68679	A human molecule a

ALIGNMENTS

RESULT 1	
AAV30337	
ID	AAV30337 standard; Protein; 1091 AA.
XX	
AC	AAV30337;
XX	
DT	15-NOV-1999 (first entry)
XX	
DE	Protein encoded by the human DIC-1 gene.
XX	
KW	DIC-1 gene; tumor suppressor gene; liver cancer; gene therapy;
KW	hepatocellular carcinoma; prostate cancer; colon cancer; rectum cancer;
KW	breast cancer; adenocarcinoma; carcinogenesis.
XX	
OS	Homo sapiens.
XX	
PN	W09943812-A2.
XX	
PD	02-SEP-1999.
XX	
PF	25-FEB-1999; 99MO-US04164.
XX	
PR	25-FEB-1998; 98US-0075952.
XX	
PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.
XX	
PI	Popescu NC, Thorngeltrsson SS, Yuan B;
XX	
DR	WPI: 1999-540590/45.

DR N-PSDB; AAC10339.

XX New nucleic acid representing the human deleted in liver cancer-1
PT gene, used for diagnosis and gene therapy of cancer

PS Claim 9; Page 54-57; 85pp; English.

CC The present sequence is encoded by the DLC-1 gene. The gene is a putative
CC tumor suppressor gene which is frequently deleted in liver cancer cells.
CC Detecting deletion of this gene, or absence of the expressed protein,
CC indicates increased susceptibility to cancer, or presence of cancer
CC (particularly hepatocellular carcinoma; cancer of prostate, colon/rectum
CC or breast, or adenocarcinoma). The DLC-1 gene, or its cDNA, can be used
CC in gene therapy to replace lost gene function, specifically for treating
CC cancer; and to generate knockout transgenic animals (as in vivo models
CC of carcinogenesis).

SO Sequence 1091 AA;

Query Match 85.7%; Score 30; DB 20; Length 1091;

Best Local Similarity 62.5%; Pred. No. 2.4e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfxy 8
11111
Db 401 wrtgsfhy 408

RESULT 2

AAB71661 AAB71661 standard; protein; 41 AA.

AC AAB71661;

DT 10-MAY-2001 (first entry)

DE Human colon associated protein #9.

KW Human; colon; cancer; disease.

OS Homo sapiens.

PN WO200112781-A1.

PD 22-FEB-2001.

PF 11-AUG-2000; 2000WO-US22157.

PR 13-AUG-1999; 99US-0148680.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Birse CE, Rosen CA;

WP1; 2001-147551/15.

CC Nucleic acids encoding 13 human colon cancer associated polypeptides,
CC useful for preventing, diagnosing and/or treating e.g. cancers
CC (especially colon cancer), Parkinson's disease and diabetic retinopathy

PS Claim 11; Page 320-321; 326pp; English.

CC The present invention relates to 13 human colon cancer-associated
CC proteins. These proteins and the nucleic acid encoding them may be
CC used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate colon cancer-associated protein
CC expression.

SO Sequence 41 AA;

Query Match 82.9%; Score 29; DB 22; Length 41;
Best Local Similarity 62.5%; Pred. No. 18;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfxy 8
11111
Db 13 wrvstfgy 20

RESULT 3

AAB53841 AAB53841 standard; Protein; 120 AA.

AC AAB53841;

DT 09-MAR-2001 (first entry)

DE Human colon cancer antigen protein sequence SEQ ID NO:1381.

KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytostatic; cardiovascular; neuroprotective; vulnerrary;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder.

OS Homo sapiens.

PN WO200055351-A1.

PD 21-SEP-2000.

PF 08-MAR-2000; 2000WO-US05883.

PR 12-MAR-1999; 99US-0124270.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;

WP1; 2000-587534/55.

DR N-PSDB; AAC98598.

CC Colon cancer associated gene sequences, referred to as colon cancer
CC antigens, useful for the treatment, prevention, and diagnosis of colon
CC disorders such as colon cancer -

PS Claim 11; Page 1952; 2104pp; English.

CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53334 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnerrary, nephrotropic, antiinfective and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present
CC invention.

SO Sequence 120 AA;

Query Match 82.9%; Score 29; DB 21; Length 120;

Best Local Similarity 62.5%; Pred. No. 48;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfyg 8
|| | |
DB 25 wrvafsfyg 32

RESULT 4
AAB07679
ID AAB07679 standard; Protein: 417 AA.
XX
AC AAB07679;
XX
DT 07-NOV-2000 (first entry)
XX
DE Amino acid sequence of Tak, a 3-oxoacyl (ACP) synthase.
XX
KW Polyketide: antibiotic Tel-Aviv; cell wall synthesis;
KM Lipid-disaccharide-pentapeptide; gingivitis.
XX
OS Myxococcus xanthus.
XX
FH Key Location/Qualifiers
FT Misc-difference 187 /note= "encoded by AGC"
FT Misc-difference 284 /note= "encoded by GGC"
FT
XX
PN EPI026248-A2.
XX
PD 09-AUG-2000.
XX
PF 31-JAN-2000; 2000EP-0300747.
XX
PR 29-JAN-1999; 99US-0240537.
XX
PA (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
XX
PI Rosenberg E, Ron E, Orr E, Paltan Y;
XX
DR WPI: 2000-500254/45.
DR N-PSDB: AAA59146.
XX
PT Novel DNA sequence involved in polyketide antibiotic Tel-Aviv
PT production useful for inhibiting cell wall synthesis and in wide range
PT of clinical applications such as treating gingivitis -
XX
PS Disclosure: Page 10; 66pp; English.
XX
CC The specification describes a DNA sequence which partially encodes
CC a functional portion of polypeptide component required for synthesizing
CC the polyketide antibiotic Tel-Aviv, postmodification of antibiotic
CC Tel-Aviv, or regulation of biosynthesis of antibiotic Tel-Aviv. The
CC antibiotic Tel-Aviv is a macrocyclic polyketide synthesised through
CC the incorporation of acetate, methionine, and glycine. It inhibits cell
CC wall synthesis by interfering with the polymerisation of the
CC lipid-disaccharide-pentapeptide. Antibiotic Tel-Aviv genes are useful
CC in combinatorial genetics, and for encoding protein components for the
CC synthesis, modification and regulation of antibiotic antibiotic Tel-Aviv.
CC Antibiotic Tel-Aviv is useful in a wide range of clinical applications
CC such as treating gingivitis. Antibiotic Tel-Aviv is also useful for
CC generating new biological agents from its secondary metabolites. The
CC present sequence represents a protein involved in synthesis of antibiotic
CC Tel-Aviv.
XX
SQ Sequence 417 AA;

Query Match 82.9%; Score 29; DB 21; Length 417;
Best Local Similarity 50.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfyg 8
|| | |

DB 76 wrsatfsq 83

RESULT 5
AAB06521
ID AAB06521 standard; peptide: 8 AA.
XX
AC AAB06521;
XX
DT 28-SEP-2000 (first entry)
XX
DE Claudin-2 cell adhesion recognition sequence SEQ ID NO: 42.
XX
KW Claudin-2 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection.
XX
OS Mammalia.
XX
PN WO200026360-A1.
XX
PD 11-MAY-2000.
XX
PF 03-NOV-1999; 99WO-CA01029.
XX
PR 03-NOV-1998; 98US-0185908.
XX
PR 30-MAR-1999; 99US-0282029.
XX
PA (ADHE-) ADHEREX TECHNOLOGIES INC.
XX
PI Blaschuck OW, Symonds JM, Gour BJ;
XX
DR WPI: 2000-365610/31.
XX
PT Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
PS Claim 46; Page 97; 121pp; English.
XX
CC The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-2 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site.
XX
SQ Sequence 8 AA;

Query Match 80.0%; Score 28; DB 21; Length 8;
Best Local Similarity 50.0%; Pred. No. 4.3e+05;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfyg 8
|| | |
DB 1 wrtsyvg 8

RESULT 6
AAB06574
ID AAB06574 standard; peptide: 8 AA.
XX
AC AAB06574;
XX
DT 28-SEP-2000 (first entry)
XX
DE Claudin-2 cyclic cell adhesion recognition sequence SEQ ID NO: 163.
XX
KW Claudin-2 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;

CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.

SO Sequence 10 AA:

Query Match 80.0%; Score 28; DB 21; Length 10;

Best Local Similarity 50.0%; Pred. No. 7.6;

Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfxx 8
11111
Db 2 wrtssyvg 9

RESULT 9

AAB06548 ID AAB06548 standard; peptide: 10 AA.

AC AAB06548;

DT 28-SEP-2000 (first entry)

DE Claudin-2 cyclic cell adhesion recognition sequence SEQ ID NO: 137.

KW Claudin-2 modulating agent; cell adhesion recognition sequence;
KM CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection; cyclic.

XX Mammalia.

PN WO200026360-A1.

PD 11-MAY-2000.

PF 03-NOV-1999; 99WO-CA01029.

PR 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.

PA (ADHE-) ADHEREX TECHNOLOGIES INC.

PI Blaschuck OW, Symonds JM, Gour BJ;

DK WPI; 2000-365610/31.

PT Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -

PS Claim 49; Page 98; 121pp; English.

CC The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-2 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.

SO Sequence 10 AA:

Query Match 80.0%; Score 28; DB 21; Length 10;

Best Local Similarity 50.0%; Pred. No. 7.6;

Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfxx 8
11111
Db 2 wrtssyvg 9

RESULT 10

AAB06557 ID AAB06557 standard; peptide: 10 AA.

AC AAB06557;

DT 28-SEP-2000 (first entry)

DE Claudin-2 cyclic cell adhesion recognition sequence SEQ ID NO: 146.

KW Claudin-2 modulating agent; cell adhesion recognition sequence;
KM CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection; cyclic.

XX Mammalia.

PN WO200026360-A1.

PD 11-MAY-2000.

PF 03-NOV-1999; 99WO-CA01029.

PR 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.

PA (ADHE-) ADHEREX TECHNOLOGIES INC.

PI Blaschuck OW, Symonds JM, Gour BJ;

DK WPI; 2000-365610/31.

PT Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -

PS Claim 49; Page 98; 121pp; English.

CC The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-2 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.

SO Sequence 10 AA:

Query Match 80.0%; Score 28; DB 21; Length 10;

Best Local Similarity 50.0%; Pred. No. 7.6;

Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfxx 8
11111
Db 2 wrtssyvg 9

RESULT 11

AAB06566 ID AAB06566 standard; peptide: 10 AA.

AC AAB06566;

DT 28-SEP-2000 (first entry)

DE Claudin-2 cyclic cell adhesion recognition sequence SEQ ID NO: 155.

XX Claudin-2 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection; cyclic.
OS Mammalia.
PN WO200026360-A1.
XX
PD 11-MAY-2000.
XX
PF 03-NOV-1999; 99WO-CA01029.
XX
PR 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
XX
PA (ADHE-) ADHEREX TECHNOLOGIES INC.
PI Blaschuck OM, Symonds JM, Gour BJ;
XX WPI: 2000-365610/31.
DR
PS Claim 49; Page 98; 121pp; English.
XX
XX The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-2 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.
XX
SO Sequence 10 AA;

Query Match 80.0%; Score 28; DB 21; Length 10;
Best Local Similarity 50.0%; Pred. No. 7.6;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
OY 1 wrxxsfxx 8
|| | | |
DB 2 wrtssyvg 9

RESULT 12
AAW20137
ID AAW20137 standard; Protein; 59 AA.
XX
AC AAW20137;
XX
DT 08-JUL-1997 (first entry)
XX
DE H. pylori cytoplasmic protein 1411681.aa.
XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
XX
OS Helicobacter pylori.
XX
PN WO9640893-A1.
XX
PD 19-DEC-1996.
XX
PF 06-JUN-1996; 96WO-US09122.
XX
PR 01-APR-1996; 96US-0630405.

PR 07-JUN-1995; 95US-0487032.
XX
PA (ASTR) ASTRA AB.
XX
PI Berglinth OF, Smith D, Mellgaard BL;
XX
DR WPI: 1997-052306/05.
DR N-PSDB; AAT67380.
XX
XX Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
XX
PS Claim 61; Page 362; 1481pp; English.
XX
XX The present sequence shows a Helicobacter pylori cytoplasmic protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
XX
SQ Sequence 59 AA;

Query Match 80.0%; Score 28; DB 18; Length 59;
Best Local Similarity 50.0%; Pred. No. 40;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 1 wrxxsfxx 8
|| | | |
DB 32 wrasgfsg 39

RESULT 13
AAV99420
ID AAV99420 standard; Protein; 205 AA.
XX
AC AAV99420;
XX
DT 08-AUG-2000 (first entry)
XX
DE Human PRO1486 (UNQ755) amino acid sequence SEQ ID NO:287.
XX
DE Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW transmembrane; secretion; immunoadhesion; pharmaceutical; screening.
XX
OS Homo sapiens.
XX
PN WO200012708-A2.
XX
PD 09-MAR-2000.
XX
PF 01-SEP-1999; 99WO-US20111.
XX
PR 01-SEP-1998; 98US-0098716.
PR 01-SEP-1998; 98US-0098749.
PR 01-SEP-1998; 98US-0098750.
PR 02-SEP-1998; 98US-0098803.
PR 02-SEP-1998; 98US-0098821.
PR 02-SEP-1998; 98US-0098843.
PR 02-SEP-1998; 98US-0099536.
PR 02-SEP-1998; 98US-0099596.
PR 09-SEP-1998; 98US-0099598.
PR 09-SEP-1998; 98US-0099602.

PR 09-SEP-1998; 98US-0099642.
PR 10-SEP-1998; 98US-0099741.
PR 10-SEP-1998; 98US-0099754.
PR 10-SEP-1998; 98US-0099763.
PR 10-SEP-1998; 98US-0099792.
PR 10-SEP-1998; 98US-0099808.
PR 10-SEP-1998; 98US-0099812.
PR 10-SEP-1998; 98US-0099815.
PR 10-SEP-1998; 98US-0099816.
PR 15-SEP-1998; 98US-0100385.
PR 15-SEP-1998; 98US-0100388.
PR 15-SEP-1998; 98US-0100390.
PR 16-SEP-1998; 98US-0100584.
PR 16-SEP-1998; 98US-0100627.
PR 16-SEP-1998; 98US-0100661.
PR 16-SEP-1998; 98US-0100662.
PR 16-SEP-1998; 98US-0100664.
PR 17-SEP-1998; 98US-0100663.
PR 17-SEP-1998; 98US-0100684.
PR 17-SEP-1998; 98US-0100710.
PR 17-SEP-1998; 98US-0100711.
PR 17-SEP-1998; 98US-0100919.
PR 17-SEP-1998; 98US-0100930.
PR 18-SEP-1998; 98US-0100848.
PR 18-SEP-1998; 98US-0100849.
PR 18-SEP-1998; 98US-0101014.
PR 18-SEP-1998; 98US-0101068.
PR 18-SEP-1998; 98US-0101071.
PR 22-SEP-1998; 98US-0101279.
PR 23-SEP-1998; 98US-0101471.
PR 23-SEP-1998; 98US-0101472.
PR 23-SEP-1998; 98US-0101474.
PR 23-SEP-1998; 98US-0101475.
PR 23-SEP-1998; 98US-0101476.
PR 23-SEP-1998; 98US-0101477.
PR 24-SEP-1998; 98US-0101479.
PR 24-SEP-1998; 98US-0101738.
PR 24-SEP-1998; 98US-0101741.
PR 24-SEP-1998; 98US-0101743.
PR 24-SEP-1998; 98US-0101915.
PR 24-SEP-1998; 98US-0101916.
PR 29-SEP-1998; 98US-0102207.
PR 29-SEP-1998; 98US-0102240.
PR 29-SEP-1998; 98US-0102307.
PR 29-SEP-1998; 98US-0102330.
PR 29-SEP-1998; 98US-0102331.
PR 30-SEP-1998; 98US-0102484.
PR 30-SEP-1998; 98US-0102487.
PR 30-SEP-1998; 98US-0102570.
PR 30-SEP-1998; 98US-0102571.
PR 01-OCT-1998; 98US-0102684.
PR 01-OCT-1998; 98US-0102687.
PR 02-OCT-1998; 98US-0102965.
PR 06-OCT-1998; 98US-0103258.
PR 06-OCT-1998; 98US-0103449.
PR 07-OCT-1998; 98US-0103314.
PR 07-OCT-1998; 98US-0103315.
PR 07-OCT-1998; 98US-0103328.
PR 07-OCT-1998; 98US-0103385.
PR 07-OCT-1998; 98US-0103396.
PR 07-OCT-1998; 98US-0103401.
PR 08-OCT-1998; 98US-0103633.
PR 08-OCT-1998; 98US-0103678.
PR 08-OCT-1998; 98US-0103679.
PR 08-OCT-1998; 98US-0103711.
PR 14-OCT-1998; 98US-0104257.
PR 20-OCT-1998; 98US-0104987.
PR 20-OCT-1998; 98US-0105000.
PR 20-OCT-1998; 98US-0105002.
PR 21-OCT-1998; 98US-0105104.
PR 22-OCT-1998; 98US-0105169.
PR 22-OCT-1998; 98US-0105266.
PR 26-OCT-1998; 98US-0105693.

PR 26-OCT-1998; 98US-0105694.
PR 27-OCT-1998; 98US-0105807.
PR 27-OCT-1998; 98US-0105881.
PR 27-OCT-1998; 98US-0105882.
PR 27-OCT-1998; 98US-0106062.
PR 28-OCT-1998; 98US-0106023.
PR 28-OCT-1998; 98US-0106029.
PR 28-OCT-1998; 98US-0106030.
PR 28-OCT-1998; 98US-0106032.
PR 28-OCT-1998; 98US-0106033.
PR 28-OCT-1998; 98US-0106178.
PR 29-OCT-1998; 98US-0106248.
PR 29-OCT-1998; 98US-0106384.
PR 29-OCT-1998; 98US-0106500.
PR 30-OCT-1998; 98US-0106464.
PR 03-NOV-1998; 98US-0106856.
PR 03-NOV-1998; 98US-0106902.
PR 03-NOV-1998; 98US-0106905.
PR 03-NOV-1998; 98US-0106919.
PR 03-NOV-1998; 98US-0106932.
PR 03-NOV-1998; 98US-0106934.
PR 10-NOV-1998; 98US-0107783.
PR 17-NOV-1998; 98US-0108775.
PR 17-NOV-1998; 98US-0108779.
PR 17-NOV-1998; 98US-0108787.
PR 17-NOV-1998; 98US-0108788.
PR 17-NOV-1998; 98US-0108801.
PR 17-NOV-1998; 98US-0108802.
PR 17-NOV-1998; 98US-0108806.
PR 17-NOV-1998; 98US-0108807.
PR 17-NOV-1998; 98US-0108867.
PR 18-NOV-1998; 98US-0108925.
PR 18-NOV-1998; 98US-0108848.
PR 18-NOV-1998; 98US-0108849.
PR 18-NOV-1998; 98US-0108850.
PR 18-NOV-1998; 98US-0108851.
PR 18-NOV-1998; 98US-0108852.
PR 18-NOV-1998; 98US-0108858.
PR 18-NOV-1998; 98US-0108904.

XX
XX (GETH) GENENTECH INC.
XX
XX Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
PI
XX WPI: 2000-237871/20.
DR N-PSDB; AAA37102.
DR
XX
XX New mammalian DNA sequences encoding transmembrane, receptor or
PT secreted PRO polypeptides, useful for screening of potential peptide or
PT small molecule inhibitors of the relevant receptor/ligand interactions
XX
PS Claim 12; Fig 162; 773pp; English.
PS
XX
XX AAA37022 to AAA37144 encode the new isolated human transmembrane,
CC receptor or secreted PRO polypeptides given in AAY9340 to AAY99462. The
CC transmembrane and receptor PRO proteins can be used for screening of
CC potential peptide or small molecule inhibitors of the relevant
CC receptor/ligand interactions. The polypeptides and nucleotide sequences
CC encoding them have various industrial applications, including uses as
CC pharmaceutical and diagnostic agents. AAA37145 to AAA3730 represent
CC PCR primers and hybridisation probes used in the isolation of the PRO
XX polypeptides from the present invention.
XX
SO Sequence 205 AA;

Query Match 80.0%; Score 28; DB 21; Length 205;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 wrxsfxg 8
I: I I I
Db 192 wksysfsg 199

PR D6-JUN-1997; 97US-0048883.
PR 06-JUN-1997; 97US-0048892.
PR 06-JUN-1997; 97US-0048895.
PR 06-JUN-1997; 97US-0048898.
PR 06-JUN-1997; 97US-0048901.
PR 06-JUN-1997; 97US-0048917.
PR 06-JUN-1997; 97US-0048963.
PR 06-JUN-1997; 97US-0048971.
PR 06-JUN-1997; 97US-0049019.
PR 06-JUN-1997; 97US-0049374.
PR 05-SEP-1997; 97US-0057627.
PR 05-SEP-1997; 97US-0057634.
PR 05-SEP-1997; 97US-0057643.
PR 05-SEP-1997; 97US-0057646.
PR 05-SEP-1997; 97US-0057649.
PR 05-SEP-1997; 97US-0057654.
PR 05-SEP-1997; 97US-0057666.
PR 05-SEP-1997; 97US-0057760.
PR 05-SEP-1997; 97US-0057763.
PR 05-SEP-1997; 97US-0057769.
PR 05-SEP-1997; 97US-0057774.
PR 05-SEP-1997; 97US-0057777.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA;
PI Fan P, Feng P, Ferlie AM, Fischer CL, Florence C;
PI Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW;
PI Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM;
PI Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;
XX WPI: 1999-059865/05.
DR N-PSDB; AAV84624.
XX
XX New isolated human genes and the secreted polypeptides they encode -
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
XX
XX Claim 11: Page 589-590; 772pp; English.
XX
XX The invention relates to nucleic acid sequences (AAV84411 to AAV84633)
CC encoding human secreted proteins (AAW88534 to AAW88756). The secreted
CC protein gene sequences are deposited with the ATCC under deposit numbers
CC ATCC 97979, 97974, 97975, 97976, 97977, 209007, 209008, 209009, 209010,
CC 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host
CC cells comprising recombinant vectors containing the nucleic acid
CC sequences are used for the recombinant production of the secreted
CC proteins. The polynucleotide and amino acid sequences are useful for are
CC useful for preventing, treating or ameliorating medical conditions e.g.
CC by protein or gene therapy. Pathological conditions can be also
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the polynucleotides, based on
CC which tissues they are most highly expressed in, and include developing
CC products for the diagnosis or treatment of cancer, neurodegenerative
CC disorders, developmental abnormalities and foetal deficiencies, blood
CC disorders, tumours, leukemias, diseases of the immune system, autoimmune
CC diseases, hepatic and renal disease, lymphomas, inflammation, allergies,
CC ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,
CC resections, prostate diseases, obesity, disorders involving osteoclasts
CC such as osteoporosis, arthritis or malignancies, diseases of testes, the
CC lung or thymus, digestive/endocrine disorders, infections and AIDS. The
CC polypeptides are also useful for identifying their binding partners.
CC The present sequence represents human secreted protein (see descriptor
CC line for gene number and clone identification).
XX
XX Sequence 206 AA:

Query Match 80.0%; Score 28; DB 20; Length 206;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxsfxx 8
| : | | |
Db 192 wkyssfsq 199
RESULT 16
AAV51676
ID AAV51676 standard; Protein: 230 AA.
XX
XX AAV51676;
AC
XX
DT 02-JUN-2000 (first entry)
XX
XX Murine clodin 2 protein.
XX
XX
XX Clodin 2; murine; tight junction-constituting membrane protein;
KW medicine.
XX
XX Mus sp.
XX
XX JP2000032984-A.
XX
PD 02-FEB-2000.
XX
PF 26-JUN-1998; 98JP-0179847.
XX
PR 15-MAY-1998; 98JP-0133215.
XX
XX (EISA) EISAI CO LTD.
XX
XX WPI: 2000-285512/25.
DR N-PSDB; AAZ89137.
XX
XX Tight junction-constituting membrane protein clodin family - useful in
PT the medical field
PT
PS Claim 2; Page 10; 22pp; Japanese.
XX
XX This invention describes novel murine nucleic acid sequences encoding the
CC clodin family of tight junction (TJ)-constituting membrane protein. The
CC membrane protein can be used in medical field. This sequence represents
CC the clodin 2 protein described in the method of the invention.
XX
SQ Sequence 230 AA:
Query Match 80.0%; Score 28; DB 21; Length 230;
Best Local Similarity 50.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 wrxxsfxx 8
| : | | |
Db 30 wrssyvg 37
RESULT 17
AAB06426
ID AAB06426 standard; peptide; 8 AA.
XX
XX AAB06426;
AC
XX
XX 28-SEP-2000 (first entry)
XX
XX Claudin-1 cell adhesion recognition sequence SEQ ID NO: 485.
DE
XX Claudin-1 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection.
XX
XX Mammalia.
XX
XX WO200026360-A1.
XX

PD 11-MAY-2000.
XX
XX 03-NOV-1999; 99WO-CA01029.
XX
PR 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
PA (ADHE-) ADHEREX TECHNOLOGIES INC.
XX
PI Blaschuck OW, Symonds JM, Gour BJ;
XX WPI: 2000-365610/31.
DR
XX Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
PS Claim 39; Page 96; 121pp; English.
CC The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-1 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site.
XX
SQ Sequence 8 AA;

Query Match 77.1%; Score 27; DB 21; Length 8;
Best Local Similarity 50.0%; Pred. No. 4.3e+05;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 wrxxsfxyg 8
11 : :
11 : :
Db 1 wrlysyag 8

RESULT 18
AAB06512
ID AAB06512 standard; peptide; 8 AA.
XX
XX AAB06512;
XX
DT 28-SEP-2000 (first entry)
XX
DE Claudin-1 cyclic cell adhesion recognition sequence SEQ ID NO: 519.
XX
XX Claudin-1 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection; cyclic.
XX
XX Mammalia.
OS
XX WO200026360-A1.
PN
XX 11-MAY-2000.
PD
XX 03-NOV-1999; 99WO-CA01029.
PE
XX 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
XX
XX (ADHE-) ADHEREX TECHNOLOGIES INC.
PA
PI Blaschuck OW, Symonds JM, Gour BJ;
XX WPI: 2000-365610/31.
DR
XX Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system

PT and across the skin -
XX
XX Claim 43; Page 97; 121pp; English.
PS
XX The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-1 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.
XX
SQ Sequence 8 AA;

Query Match 77.1%; Score 27; DB 21; Length 8;
Best Local Similarity 50.0%; Pred. No. 4.3e+05;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 wrxxsfxyg 8
11 : :
11 : :
Db 1 wrlysyag 8

RESULT 19
AAB06583
ID AAB06583 standard; peptide; 8 AA.
XX
XX AAB06583;
XX
DT 28-SEP-2000 (first entry)
XX
DE Claudin-3 cell adhesion recognition sequence SEQ ID NO: 58.
XX
XX Claudin-3 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection.
XX
XX Mammalia.
OS
XX WO200026360-A1.
PN
XX 11-MAY-2000.
PD
XX 03-NOV-1999; 99WO-CA01029.
PE
XX 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
XX
XX (ADHE-) ADHEREX TECHNOLOGIES INC.
PA
PI Blaschuck OW, Symonds JM, Gour BJ;
XX WPI: 2000-365610/31.
DR
XX Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
PS Claim 52; Page 99; 121pp; English.
CC The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-3 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site.
XX
SQ Sequence 8 AA;

Query Match 77.1%; Score 27; DB 21; Length 8;
Best Local Similarity 50.0%; Pred. No. 4.3e+05;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxsfxg 8
11:11
Db 1 wrvsalfg 8

Result 20

AAB06636
ID AAB06636 standard; peptide; 8 AA.

AC AAB06636;

DT 28-SEP-2000 (first entry)

DE Claudin-3 cyclic cell adhesion recognition sequence SEQ ID NO: 269.

KW Claudin-3 modulating agent; cell adhesion recognition sequence;

KW CAR sequence; autoimmune disease; inflammatory disease; cancer;

KW graft rejection; cyclic.

OS Mammalia.

PN WO200026360-A1.

PD 11-MAY-2000.

PF 03-NOV-1999; 99WO-CA01029.

PR 03-NOV-1998; 98US-0185908.

PR 30-MAR-1999; 99US-0282029.

PA (ADHE-) ADHEREX TECHNOLOGIES INC.

PI Blaschuck OM, Symonds JM, Gour BJ;

WPI: 2000-365610/31.

PS Claim 55; Page 100; 121pp; English.

CC The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-3 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.

SO Sequence 8 AA;

Query Match 77.1%; Score 27; DB 21; Length 8;
Best Local Similarity 50.0%; Pred. No. 4.3e+05;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxsfxg 8
11:11
Db 1 wrvsalfg 8

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:23:44 : Search time 48.05 Seconds
(without alignments)
3.747 Million cell updates/sec

Title: 09-185908-1g
Perfect score: 35
Sequence: 1 wrxxsfxy 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues
Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

- Database : Issued_Patents_AA:*
- 1: /cgn2_6/ptodata/2/1aa/5A-COMB.pep:*
 - 2: /cgn2_6/ptodata/2/1aa/5B-COMB.pep:*
 - 3: /cgn2_6/ptodata/2/1aa/6A-COMB.pep:*
 - 4: /cgn2_6/ptodata/2/1aa/6B-COMB.pep:*
 - 5: /cgn2_6/ptodata/2/1aa/PCtus-COMB.pep:*
 - 6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	77.1	345	3 US-09-222-817-2	Sequence 2, Appl
2	27	77.1	345	4 US-09-222-786-2	Sequence 2, Appl
3	27	77.1	530	3 US-09-222-817-12	Sequence 12, Appl
4	27	77.1	530	3 US-09-222-817-14	Sequence 14, Appl
5	27	77.1	530	4 US-09-222-786-12	Sequence 12, Appl
6	27	77.1	530	4 US-09-222-786-14	Sequence 14, Appl
7	26	74.3	462	4 US-09-352-159-6	Sequence 6, Appl
8	26	74.3	462	4 US-09-352-159-8	Sequence 8, Appl
9	26	74.3	462	4 US-09-352-168-6	Sequence 6, Appl
10	26	74.3	462	4 US-09-352-168-8	Sequence 8, Appl
11	26	74.3	463	4 US-09-352-159-11	Sequence 11, Appl
12	26	74.3	463	4 US-09-352-168-11	Sequence 11, Appl
13	26	74.3	487	4 US-09-352-159-21	Sequence 21, Appl
14	26	74.3	487	4 US-09-352-168-21	Sequence 21, Appl
15	26	74.3	554	4 US-09-352-159-17	Sequence 17, Appl
16	26	74.3	554	4 US-09-352-168-17	Sequence 17, Appl
17	26	74.3	591	4 US-09-352-159-44	Sequence 44, Appl
18	26	74.3	591	4 US-09-352-159-46	Sequence 46, Appl
19	26	74.3	598	4 US-09-352-159-40	Sequence 40, Appl
20	26	74.3	598	4 US-09-352-159-42	Sequence 42, Appl
21	26	74.3	600	4 US-09-352-159-23	Sequence 23, Appl
22	26	74.3	600	4 US-09-352-159-36	Sequence 36, Appl
23	26	74.3	600	4 US-09-352-159-38	Sequence 38, Appl
24	26	74.3	600	4 US-09-352-168-23	Sequence 23, Appl
25	26	74.3	692	4 US-09-352-159-19	Sequence 19, Appl
26	26	74.3	692	4 US-09-352-168-19	Sequence 19, Appl
27	26	74.3	829	4 US-09-352-159-33	Sequence 33, Appl

28	26	74.3	829	4 US-09-352-168-33	Sequence 33, Appl
29	26	74.3	991	4 US-09-352-159-27	Sequence 27, Appl
30	26	74.3	991	4 US-09-352-168-27	Sequence 27, Appl
31	26	74.3	1000	4 US-09-352-159-25	Sequence 25, Appl
32	26	74.3	1000	4 US-09-352-168-25	Sequence 25, Appl
33	26	74.3	1196	4 US-09-352-159-31	Sequence 31, Appl
34	26	74.3	1196	4 US-09-352-168-31	Sequence 31, Appl
35	26	74.3	1205	4 US-09-352-159-29	Sequence 29, Appl
36	26	74.3	1205	4 US-09-352-168-29	Sequence 29, Appl
37	25	71.4	34	2 US-08-751-767A-38	Sequence 38, Appl
38	25	71.4	154	1 US-08-153-848A-36	Sequence 36, Appl
39	25	71.4	154	3 US-09-299-843A-36	Sequence 36, Appl
40	25	71.4	154	5 PCW-US93-11153-36	Sequence 36, Appl
41	25	71.4	222	4 US-09-140-804-7	Sequence 7, Appl
42	25	71.4	360	4 US-09-116-498-8	Sequence 8, Appl
43	25	71.4	360	4 US-09-116-498-10	Sequence 10, Appl
44	25	71.4	360	4 US-09-116-498-12	Sequence 12, Appl
45	25	71.4	2186	2 US-08-822-445-2	Sequence 2, Appl
46	25	71.4	2476	2 US-08-276-967-2	Sequence 2, Appl
47	25	71.4	3672	2 US-08-822-445-12	Sequence 12, Appl
48	25	71.4	3801	2 US-08-822-445-10	Sequence 10, Appl
49	25	71.4	5405	4 US-08-718-388-9	Sequence 9, Appl
50	24	68.6	9	1 US-08-660-626-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-09-222-817-2
: Sequence 2, Application US/09222817
: Patent No. 6037154
: GENERAL INFORMATION:
: APPLICANT: MIKIKO SUGA, Masakazu SUGIMOTO, Tsuyoshi OSUMI, Tsuyoshi NAKAMATSU,
: TITLE OF INVENTION: METHOD OF PRODUCING L-SERINE BY FERMENTATION
: FILE REFERENCE: OP813
: CURRENT APPLICATION NUMBER: US/09/222, 817
: EARLIER FILING DATE: 1998-12-30
: EARLIER APPLICATION NUMBER: JP 10-3751
: EARLIER FILING DATE: 1998-01-12
: EARLIER APPLICATION NUMBER: JP 10-353521
: EARLIER FILING DATE: 1998-12-11
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 2
: LENGTH: 345
: TYPE: PRT
: ORGANISM: Corynebacterium glutamicum
US-09-222-817-2

Query Match 77.1%; Score 27; DB 3; Length 345;
Best Local Similarity 50.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxsfxy 8
I: | | |
Db 132 WKRSSFNG 139
RESULT 2
US-09-222-786-2
: Sequence 2, Application US/09222786A
: Patent No. 6258573
: GENERAL INFORMATION:
: APPLICANT: MIKIKO SUGA, Masakazu SUGIMOTO, Tsuyoshi OSUMI, Tsuyoshi NAKAMATSU,
: TITLE OF INVENTION: METHOD OF PRODUCING L-SERINE BY FERMENTATION
: FILE REFERENCE: OP812
: CURRENT APPLICATION NUMBER: US/09/222, 786A
: EARLIER FILING DATE: 1998-12-30
: EARLIER APPLICATION NUMBER: JP 10-3751
: EARLIER FILING DATE: 1998-01-12
: EARLIER APPLICATION NUMBER: JP 10-353513

EARLIER FILING DATE: 1998-12-11
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 2
LENGTH: 345
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-222-786-2

Query Match 77.1%; Score 27; DB 4; Length 345;
Best Local Similarity 50.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfkg 8
|: || |
Db 132 WKRSSFNG 139

RESULT 3
US-09-222-817-12
Sequence 12, Application US/09222817
Patent No. 6037154
GENERAL INFORMATION:
APPLICANT: MIKIKO SUGA, Masakazu SUGIMOTO, Tsuyoshi OSUMI, Tsuyoshi NAKAMATSU,
FILE REFERENCE: OP813
CURRENT APPLICATION NUMBER: US/09/222,817
EARLIER FILING DATE: 1998-12-30
EARLIER APPLICATION NUMBER: JP 10-3751
EARLIER FILING DATE: 1998-01-12
EARLIER APPLICATION NUMBER: JP 10-353521
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 12
LENGTH: 530
TYPE: PRT
ORGANISM: Brevibacterium flavum
US-09-222-817-12

Query Match 77.1%; Score 27; DB 3; Length 530;
Best Local Similarity 50.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfkg 8
|: || |
Db 132 WKRSSFNG 139

RESULT 4
US-09-222-817-14
Sequence 14, Application US/09222817
Patent No. 6037154
GENERAL INFORMATION:
APPLICANT: MIKIKO SUGA, Masakazu SUGIMOTO, Tsuyoshi OSUMI, Tsuyoshi NAKAMATSU,
FILE REFERENCE: OP813
CURRENT APPLICATION NUMBER: US/09/222,817
EARLIER FILING DATE: 1998-12-30
EARLIER APPLICATION NUMBER: JP 10-3751
EARLIER FILING DATE: 1998-01-12
EARLIER APPLICATION NUMBER: JP 10-353521
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 14
LENGTH: 530
TYPE: PRT
ORGANISM: Brevibacterium flavum
US-09-222-817-14

Query Match 77.1%; Score 27; DB 3; Length 530;
Best Local Similarity 50.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfkg 8
|: || |
Db 132 WKRSSFNG 139

RESULT 5
US-09-222-786-12
Sequence 12, Application US/09222786A
Patent No. 6258573
GENERAL INFORMATION:
APPLICANT: MIKIKO SUGA, Masakazu SUGIMOTO, Tsuyoshi OSUMI, Tsuyoshi NAKAMATSU,
FILE REFERENCE: OP812
CURRENT APPLICATION NUMBER: US/09/222,786A
EARLIER FILING DATE: 1998-12-30
EARLIER APPLICATION NUMBER: JP 10-3751
EARLIER FILING DATE: 1998-01-12
EARLIER APPLICATION NUMBER: JP 10-353513
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 12
LENGTH: 530
TYPE: PRT
ORGANISM: Brevibacterium flavum
US-09-222-786-12

Query Match 77.1%; Score 27; DB 4; Length 530;
Best Local Similarity 50.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfkg 8
|: || |
Db 132 WKRSSFNG 139

RESULT 6
US-09-222-786-14
Sequence 14, Application US/09222786A
Patent No. 6258573
GENERAL INFORMATION:
APPLICANT: MIKIKO SUGA, Masakazu SUGIMOTO, Tsuyoshi OSUMI, Tsuyoshi NAKAMATSU,
FILE REFERENCE: OP812
CURRENT APPLICATION NUMBER: US/09/222,786A
EARLIER FILING DATE: 1998-12-30
EARLIER APPLICATION NUMBER: JP 10-3751
EARLIER FILING DATE: 1998-01-12
EARLIER APPLICATION NUMBER: JP 10-353513
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 14
LENGTH: 530
TYPE: PRT
ORGANISM: Brevibacterium flavum
US-09-222-786-14

Query Match 77.1%; Score 27; DB 4; Length 530;
Best Local Similarity 50.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxsfkg 8
|: || |
Db 132 WKRSSFNG 139

RESULT 7
US-09-352-159-6
; Sequence 6, Application US/09352159A
; Patent No. 6211434
; GENERAL INFORMATION:
; APPLICANT: Duivick, Jonathan P.
; APPLICANT: Gilliam, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; TITLE OF INVENTION: Polynucleotides and Related Polypeptides and Methods of Use
; FILE REFERENCE: 1134
; CURRENT APPLICATION NUMBER: US/09/352,159A
; CURRENT FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; EARLIER APPLICATION NUMBER: 60/135,391
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 462
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-352-159-6

Query Match
Best Local Similarity 74.3%; Score 26; DB 4; Length 462;
Best Local Similarity 50.0%; Pred. No. 2.6e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxxsfxxg 8
|||
DB 310 WREGFGSG 317

RESULT 8
US-09-352-159-8
; Sequence 8, Application US/09352159A
; Patent No. 6211434
; GENERAL INFORMATION:
; APPLICANT: Duivick, Jonathan P.
; APPLICANT: Gilliam, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; TITLE OF INVENTION: Polynucleotides and Related Polypeptides and Methods of Use
; FILE REFERENCE: 1134
; CURRENT APPLICATION NUMBER: US/09/352,159A
; CURRENT FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; EARLIER APPLICATION NUMBER: 60/135,391
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 462
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-352-159-8

Query Match
Best Local Similarity 74.3%; Score 26; DB 4; Length 462;
Best Local Similarity 50.0%; Pred. No. 2.6e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxxsfxxg 8
|||
DB 310 WREGFGSG 317

RESULT 9
US-09-352-168-6

; Sequence 6, Application US/09352168A
; Patent No. 6211435
; GENERAL INFORMATION:
; APPLICANT: Crasta, Oswald R.
; APPLICANT: Duivick, Jonathan P.
; APPLICANT: Folkerts, Otto
; APPLICANT: Gilliam, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; TITLE OF INVENTION: Polynucleotides and Related Polypeptides and Methods of Use
; FILE REFERENCE: 0875
; CURRENT APPLICATION NUMBER: US/09/352,168A
; CURRENT FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 462
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-352-168-6

Query Match
Best Local Similarity 74.3%; Score 26; DB 4; Length 462;
Best Local Similarity 50.0%; Pred. No. 2.6e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxxsfxxg 8
|||
DB 310 WREGFGSG 317

RESULT 10
US-09-352-168-8
; Sequence 8, Application US/09352168A
; Patent No. 6211435
; GENERAL INFORMATION:
; APPLICANT: Crasta, Oswald R.
; APPLICANT: Duivick, Jonathan P.
; APPLICANT: Folkerts, Otto
; APPLICANT: Gilliam, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; TITLE OF INVENTION: Polynucleotides and Related Polypeptides and Methods of Use
; FILE REFERENCE: 0875
; CURRENT APPLICATION NUMBER: US/09/352,168A
; CURRENT FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 462
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-352-168-8

Query Match
Best Local Similarity 74.3%; Score 26; DB 4; Length 462;
Best Local Similarity 50.0%; Pred. No. 2.6e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxxsfxxg 8
|||
DB 310 WREGFGSG 317

RESULT 11
US-09-352-159-11
; Sequence 11, Application US/09352159A
; Patent No. 6211434
; GENERAL INFORMATION:

```
; APPLICANT: Duvick, Jonathan P.
; APPLICANT: Giliham, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; FILE REFERENCE: 1134
; CURRENT APPLICATION NUMBER: US/09/352,159A
; EARLIER FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; EARLIER APPLICATION NUMBER: 60/135,391
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-352-159-11
```

```
Query Match          74.3%; Score 26; DB 4; Length 463;
Best Local Similarity 50.0%; Pred. No. 2.6e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
OY 1 wxsfxg 8
|| | |
Db 311 WREGFGSG 318
```

```
RESULT 12
US-09-352-168-11
; Sequence 11, Application US/09352168A
; Patent No. 6211435
; GENERAL INFORMATION:
```

```
; APPLICANT: Crasta, Oswald R.
; APPLICANT: Duvick, Jonathan P.
; APPLICANT: Folkerts, Otto
; APPLICANT: Giliham, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; TITLE OF INVENTION: Polynucleotides and Related Polypeptides and Methods of Use
; FILE REFERENCE: 0875
; CURRENT APPLICATION NUMBER: US/09/352,168A
; CURRENT FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: Extra lysine in the polypeptide sequence of
; OTHER INFORMATION: K:TRAPAO, 463 aa.
US-09-352-168-11
```

```
Query Match          74.3%; Score 26; DB 4; Length 463;
Best Local Similarity 50.0%; Pred. No. 2.6e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
OY 1 wxsfxg 8
|| | |
Db 311 WREGFGSG 318
```

```
RESULT 13
US-09-352-159-21
; Sequence 21, Application US/09352159A
; Patent No. 6211434
; GENERAL INFORMATION:
```

```
; APPLICANT: Duvick, Jonathan P.
; APPLICANT: Giliham, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; FILE REFERENCE: 1134
; CURRENT APPLICATION NUMBER: US/09/352,159A
; EARLIER FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; EARLIER APPLICATION NUMBER: 60/135,391
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(24)
US-09-352-159-21
```

```
Query Match          74.3%; Score 26; DB 4; Length 487;
Best Local Similarity 50.0%; Pred. No. 2.7e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
OY 1 wxsfxg 8
|| | |
Db 335 WREGFGSG 342
```

```
RESULT 14
US-09-352-168-21
; Sequence 21, Application US/09352168A
; Patent No. 6211435
; GENERAL INFORMATION:
```

```
; APPLICANT: Crasta, Oswald R.
; APPLICANT: Duvick, Jonathan P.
; APPLICANT: Folkerts, Otto
; APPLICANT: Giliham, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; TITLE OF INVENTION: Polynucleotides and Related Polypeptides and Methods of Use
; FILE REFERENCE: 0875
; CURRENT APPLICATION NUMBER: US/09/352,168A
; CURRENT FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(24)
; OTHER INFORMATION: K:TRAPAO translational fusion with barley alpha
; OTHER INFORMATION: amylase signal sequence, for expression and
; OTHER INFORMATION: secretion of the mature TRAPAO in maize.
US-09-352-168-21
```

```
Query Match          74.3%; Score 26; DB 4; Length 487;
Best Local Similarity 50.0%; Pred. No. 2.7e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
OY 1 wxsfxg 8
|| | |
Db 335 WREGFGSG 342
```

RESULT 15
US-09-352-159-17
; Sequence 17, Application US/09352159A
; Patent No. 6211434
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jonathan P.
; APPLICANT: Gilliam, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; FILE REFERENCE: 1134
; CURRENT APPLICATION NUMBER: US/09/352,159A
; EARLIER FILING DATE: 1998-07-12
; EARLIER FILING DATE: 1998-07-25
; EARLIER APPLICATION NUMBER: 60/135,391
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 554
; TYPE: PRT
; ORGANISM: Exophiala spiniifera
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(89)
US-09-352-159-17

Query Match 74.3%; Score 26; DB 4; Length 554;
Best Local Similarity 50.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxxsfxx 8
|||
DB 402 WREGGFG 409

RESULT 16
US-09-352-168-17
; Sequence 17, Application US/09352168A
; Patent No. 6211435
; GENERAL INFORMATION:
; APPLICANT: Cresta, Oswald R.
; APPLICANT: Duvick, Jonathan P.
; APPLICANT: Folkerts, Otto
; APPLICANT: Gilliam, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; FILE REFERENCE: 0875
; CURRENT APPLICATION NUMBER: US/09/352,168A
; EARLIER FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 554
; TYPE: PRT
; ORGANISM: Exophiala spiniifera
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(89)
; OTHER INFORMATION: yeast alpha mating factor secretion signal.
US-09-352-168-17

Query Match 74.3%; Score 26; DB 4; Length 554;
Best Local Similarity 50.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxxsfxx 8
|||
DB 402 WREGGFG 409

RESULT 17
US-09-352-159-44
; Sequence 44, Application US/09352159A
; Patent No. 6211434
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jonathan P.
; APPLICANT: Gilliam, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; FILE REFERENCE: 1134
; CURRENT APPLICATION NUMBER: US/09/352,159A
; EARLIER FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; EARLIER APPLICATION NUMBER: 60/135,391
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 44
; LENGTH: 591
; TYPE: PRT
; ORGANISM: Rhinocladella atrovirens
US-09-352-159-44

Query Match 74.3%; Score 26; DB 4; Length 591;
Best Local Similarity 50.0%; Pred. No. 3.3e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxxsfxx 8
|||
DB 439 WREGGFG 446

RESULT 18
US-09-352-159-46
; Sequence 46, Application US/09352159A
; Patent No. 6211434
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jonathan P.
; APPLICANT: Folkerts, Otto
; APPLICANT: Gilliam, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; FILE REFERENCE: 1134
; CURRENT APPLICATION NUMBER: US/09/352,159A
; EARLIER FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; EARLIER APPLICATION NUMBER: 60/135,391
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 46
; LENGTH: 591
; TYPE: PRT
; ORGANISM: Rhinocladella atrovirens
US-09-352-159-46

Query Match 74.3%; Score 26; DB 4; Length 591;
Best Local Similarity 50.0%; Pred. No. 3.3e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxxsfxx 8
|||
DB 442 WREGGFG 449

RESULT 19
US-09-352-159-40

; Sequence 40, Application US/09352159A
; Patent No. 6211434
; GENERAL INFORMATION:
; APPLICANT: DUVICK, Jonathan P.
; APPLICANT: GILLIAM, Jacob T.
; APPLICANT: MADDOX, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; TITLE OF INVENTION: Polynucleotides and Related Polypeptides and Methods of Use
; FILE REFERENCE: 1134
; CURRENT APPLICATION NUMBER: US/09/352,159A
; EARLIER FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; EARLIER APPLICATION NUMBER: 60/135,391
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 40
; LENGTH: 598
; TYPE: PRT
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (216)...(216)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-352-159-40

Query Match 74.3%; Score 26; DB 4; Length 598;
Best Local Similarity 50.0%; Pred. No. 3.3e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 wrxsfxg 8
Db 446 WREQGFSG 453

RESULT 20
US-09-352-159-42

; Sequence 42, Application US/09352159A
; Patent No. 6211434
; GENERAL INFORMATION:
; APPLICANT: DUVICK, Jonathan P.
; APPLICANT: GILLIAM, Jacob T.
; APPLICANT: MADDOX, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; TITLE OF INVENTION: Polynucleotides and Related Polypeptides and Methods of Use
; FILE REFERENCE: 1134
; CURRENT APPLICATION NUMBER: US/09/352,159A
; EARLIER FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; EARLIER APPLICATION NUMBER: 60/135,391
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 42
; LENGTH: 598
; TYPE: PRT
; ORGANISM: Rhinocladiaella atrovirens
US-09-352-159-42

Db 446 WREQGFSG 453

Search completed: January 14, 2002, 07:23:45
Job time: 76 sec

Query Match 74.3%; Score 26; DB 4; Length 598;
Best Local Similarity 50.0%; Pred. No. 3.3e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 wrxsfxg 8
|| | |

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:37:32 ; Search time 63.57 Seconds
(without alignments)
9.586 Million cell updates/sec

Title: 09-185908-1h
Perfect score: 35
Sequence: 1 wrxxafxg 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : PIR_68: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	91.4	199	1	TUBPKL
2	32	91.4	224	2	B85584
3	30	85.7	159	2	S41176
4	30	85.7	280	2	A39484
5	30	85.7	330	2	T34972
6	30	85.7	370	2	B83191
7	29	82.9	198	2	G85506
8	29	82.9	379	2	D70786
9	29	82.9	937	2	S78561
10	28	80.0	206	2	H75258
11	28	80.0	213	2	A56152
12	28	80.0	380	2	T40965
13	28	80.0	230	2	T02995
14	27	77.1	224	2	G85842
15	27	77.1	293	2	E85718
16	27	77.1	293	2	A85817
17	27	77.1	369	1	D64763
18	27	77.1	369	1	S57525
19	27	77.1	369	2	G85530
20	27	77.1	373	1	A33419
21	27	77.1	373	1	DEPTA
22	27	77.1	373	1	S68061
23	27	77.1	374	1	DEHUC2
24	27	77.1	374	1	A56643
25	27	77.1	376	1	JC4967
26	27	77.1	376	1	S51187
27	27	77.1	378	1	A49662
28	27	77.1	378	1	H64052
29	27	77.1	378	2	F81097

30	27	77.1	379	1	S51357	alcohol dehydrogen
31	27	77.1	379	1	S71244	alcohol dehydrogen
32	27	77.1	381	1	UN0447	alcohol dehydrogen
33	27	77.1	381	2	T03289	formaldehyde dehyd
34	27	77.1	381	2	T04164	alcohol dehydrogen
35	27	77.1	386	1	S31140	alcohol dehydrogen
36	27	77.1	396	2	S31059	acyl-lacyl-carrier
37	27	77.1	513	1	A35742	acetyl-in (EC 3.4.
38	27	77.1	733	1	S33643	transforming prote
39	27	77.1	1083	2	S54293	regulator protein
40	26	74.3	154	2	E70971	probable RNA meth
41	26	74.3	169	2	B71454	hypothetical prote
42	26	74.3	171	2	H75017	hypothetical prote
43	26	74.3	212	2	S72873	hypothetical prote
44	26	74.3	224	2	A85742	probable tail comp
45	26	74.3	232	2	E70830	hypothetical prote
46	26	74.3	245	2	C85682	hypothetical prote
47	26	74.3	260	2	F82120	zinc ABC transport
48	26	74.3	261	2	A64066	probable membrane
49	26	74.3	262	2	F82959	permease of ABC z1
50	26	74.3	286	2	D64235	hypothetical prote

ALIGNMENTS

RESULT 1
TUBPKL
tail assembly protein K - phage lambda
C:Species: phage lambda
C:Date: 13-Jun-1983 #sequence-revision 13-Jun-1983 #text-change 23-Jul-1999
C:Accession: H43009; G43013; A04355
R.Daniels, D.
submitted to the Nucleic Acid Sequence Database, September 1982
A:Reference number: A94614
A:Accession: H43009
A:Molecule type: DNA
A:Residues: 1-199 <DAN>
R.Sanger, F.; Coulson, A.R.; Hong, G.F.; Hill, D.F.; Petersen, G.B.
J. Mol. Biol. 162, 729-773, 1982
A:Title: Nucleotide sequence of bacteriophage lambda DNA.
A:Reference number: A92891; M01D:83189071
A:Accession: G43013
A:Molecule type: DNA
A:Residues: 1-199 <SAN>
A:Cross-references: GB:J02459; GB:M17233; GB:M24325; GB:V00636; GB:X00906; NID:921510
A:Note: there are two possible initiation sites for gene K translation, the codon for C:Comment: Gene K protein is involved in the assembly of the initiator complex for ta
C:Genetics:
A:Gene: K
A:Map position: 29.43-30.66
C:Superfamily: phage lambda tail assembly protein K

Query Match 91.4%; Score 32; DB 1; Length 199;
Best Local Similarity 62.5%; Pred. No. 3.3;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
DB 180 WRASAFRG 187

RESULT 2
B85584
probable tail component of prophage CP-933K Z0978 [Imported] - Escherichia coli (stra
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence-revision 16-Feb-2001 #text-change 31-Mar-2001
C:Accession: B85584
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
hiller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Diallanla, E.; Potamoustis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: B85584
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-224 <STO>
A:Cross-references: GB:AE005174; NID:912513746; PIDN:AAG55134.1; GSPDB:GN00145; UWCP:Z05
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: 20978
C:Superfamily: phage lambda tail assembly protein K

Query Match 91.4%; Score 32; DB 2; Length 224;
Best Local Similarity 62.5%; Pred. No. 3.6;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
DB 205 WRSAFTG 212

RESULT 3
S41178
gene 36 protein - phage SPPI
C:Species: phage SPPI
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 11-May-2000
C:Accession: S43804; T42337; S41178
R:Pedre, X.; Weise, F.; Chai, S.; Lueder, G.; Alonso, J.C.
J. Mol. Biol. 236, 1324-1340, 1994
A:Title: Analysis of cis and trans acting elements required for the initiation of DNA re
A:Reference number: S43798; MUID:94172631
A:Accession: S43804
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-159 <PE2>
A:Cross-references: EMBL:X67865; NID:9472886; PIDN:CAA48055.1; PID:9439635
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, July 1992
R:Alonso, J.C.; Luder, G.; Stiege, A.C.; Chai, S.; Weise, F.; Trautner, T.A.
Gene 204, 201-212, 1997
A:Title: The complete nucleotide sequence and functional organization of Bacillus subtil
A:Reference number: Z22137; MUID:98094274
A:Accession: T42337
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-159 <ALO>
A:Cross-references: EMBL:X97918; PIDN:CAA66491.1
C:Superfamily: single-stranded DNA-binding protein; single-stranded DNA-binding protein
F:17-93/Domain: single-stranded DNA-binding protein homology <SSD>

Query Match 85.7%; Score 30; DB 2; Length 159;
Best Local Similarity 62.5%; Pred. No. 7.6;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
DB 77 WRTGAFEG 84

RESULT 4
A39484
androgen-withdrawal apoptosis protein RVP1, prostatic - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 03-Aug-1992 #sequence_revision 03-Aug-1992 #text_change 04-Mar-2000
C:Accession: A39484
R:Biehl, M.M.; Miesfeld, R.L.
Mol. Endocrinol. 5, 1381-1388, 1991
A:Title: Isolation and characterization of transcripts induced by androgen withdrawal an
A:Reference number: A39484; MUID:92130987
A:Accession: A39484
A:Molecule type: mRNA
A:Residues: 1-280 <BRI>
A:Cross-references: GB:M74067; NID:9205857; PIDN:AAA41760.1; PID:9205858

C:Genetics:
A:Gene: RVP.1
C:Superfamily: rat androgen-withdrawal apoptosis protein RVP1

Query Match 85.7%; Score 30; DB 2; Length 280;
Best Local Similarity 62.5%; Pred. No. 13;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
DB 29 WRVSAFIG 36

RESULT 5
T34972
Probable membrane protein - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C:Accession: T34972
R:Saunders, D.C.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M
submitted to the EMBL Data Library, August 1999
A:Reference number: Z21563
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-330 <SAU>
A:Cross-references: EMBL:AL109663; PIDN:CAB52011.1; GSPDB:GN00070; SCOEDB:SC4A10.35C
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC4A10.35C

Query Match 85.7%; Score 30; DB 2; Length 330;
Best Local Similarity 62.5%; Pred. No. 15;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
DB 31 WRLAFAG 38

RESULT 6
B83191
alcohol dehydrogenase class III PA3629 [imported] - Pseudomonas aeruginosa (strain PA
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: B83191
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: AB2950; MUID:20437337
A:Accession: B83191
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-370 <STO>
A:Cross-references: GB:AE004783; GB:AE004091; NID:99949786; PIDN:AAG07017.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: adhC; PA3629
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 85.7%; Score 30; DB 2; Length 370;
Best Local Similarity 62.5%; Pred. No. 17;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
DB 310 WRGSAFIG 317

RESULT 7	8855506	hypothetical protein 20246 [Imported] - Escherichia coli (strain O157:H7)
C:Species:	Escherichia coli	
C:Date:	16-Feb-2001	#sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession:	GB55506	
R:Peena, N.T.;	Plunkett III, G.;	Burland, V.;
Miller, L.;	Grobeck, E.J.;	Davls, N.W.;
Nature 409,	529-533,	2001
A:Title:	Genome sequence of enterohemorrhagic Escherichia coli O157:H7.	
A:Reference number:	AB5480;	MUID:21074935; PMID:11206551
A:Accession:	GB5506	
A:Status:	Preliminary	
A:Molecule type:	DNA	
A:Residues:	1-198	<STG>
A:Cross-references:	GB:AE005174;	NID:g12512949; PIDN:AAG54515.1; GSPDB:GN00145; UMGCP:Z02
A:Experimental source:	strain O157:H7,	substrain EDL933
C:Genetics:		
A:Gene:	Z0246	
Query Match	Best Local Similarity	82.9%; Score 29; DB 2; Length 198;
Matches 5;	Conservative 0;	Mismatches 3; Indels 0; Gaps 0;
OY 1	wrxaxfxg 8	
Db 48	WRKRAFNG 55	
RESULT 8	D70786	probable gcvT protein - Mycobacterium tuberculosis (strain H37RV)
C:Species:	Mycobacterium tuberculosis	
C:Date:	17-Jul-1998	#sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession:	D70786	
R:Coie, S.T.;	Broesch, R.;	Parkhill, J.;
Connor, R.;	Davies, R.;	Devlin, K.;
Rajandream, M.A.;	Rogers, J.;	Rutler, S.;
Nature 393,	537-544,	1998
A:Authors:	Squares, R.;	Sulston, J.E.;
A:Title:	Deciphering the biology of Mycobacterium tuberculosis from the complete genome	
A:Reference number:	AV0500;	MUID:98295987
A:Accession:	D70786	
A:Status:	Preliminary	
A:Molecule type:	DNA	
A:Residues:	1-379	<COL>
A:Cross-references:	GB:Z70283;	GB:AL123456; NID:g3261561; PIDN:CAA94254.1; PID:g12337064
A:Experimental source:	strain H37RV	
C:Genetics:		
A:Gene:	gcvT	
C:Superfamily:	amlnomethyltransferase	
Query Match	Best Local Similarity	82.9%; Score 29; DB 2; Length 379;
Matches 5;	Conservative 0;	Mismatches 3; Indels 0; Gaps 0;
OY 1	wrxaxfxg 8	
Db 275	WRKDAFG 282	
RESULT 9	S78561	CS3 pilin synthesis protein, 104K - Escherichia coli
N:Contains:	CS3 pilin synthesis protein, 104K;	CS3 pilin synthesis protein, 20K; CS3 pil
C:Species:	Escherichia coli	
C:Date:	13-Sep-1998	#sequence_revision 13-Sep-1998 #text_change 21-Jul-2000
C:Accession:	S78561;	S07900;
R:Jalajukumar, M.B.;	Thomas, C.J.;	Halter, R.;
Mol. Microbiol.	3,	1685-1695,

A:Reference number: S07899; MUID:90158116	
A:Accession: S78561	
A:Molecule type: DNA	
A:Residues: 1-937 <JAL>	
A:Cross-references: EMBL:X16944	
A:Experimental source: strain PB176	
A:Note: the readthrough stop codon TAG for residue 754 is translated as X	
A:Accession: S07900	
A:Molecule type: DNA	
A:Residues: 181-753 <JAM>	
A:Cross-references: EMBL:X16944; NID:g41155; PIDN:CAA34816.1; PID:g41157	
A:Experimental source: strain PB176	
A:Accession: S07902	
A:Molecule type: DNA	
A:Residues: 451-753 <JAA>	
A:Cross-references: EMBL:X16944; NID:g41155; PIDN:CAA34818.1; PID:g41159	
A:Experimental source: strain PB176	
A:Accession: S07903	
A:Molecule type: DNA	
A:Residues: 572-753 <JAZ>	
A:Cross-references: EMBL:X16944; NID:g41155; PIDN:CAA34819.1; PID:g41160	
A:Experimental source: strain PB176	
C:Function:	
A:Description: Involved in the biosynthesis of CS3 pilin	
F:1-937/Product: CS3 pilin synthesis protein, 104K #status predicted <MA1>	
F:181-753/Product: CS3 pilin synthesis protein, 63K #status predicted <MA2>	
F:317-753/Product: CS3 pilin synthesis protein, 48K #status predicted <MA3>	
F:451-753/Product: CS3 pilin synthesis protein, 33K #status predicted <MA4>	
F:572-753/Product: CS3 pilin synthesis protein, 20K #status predicted <MA5>	
Query Match	82.9%; Score 29; DB 2; Length 937;
Best Local Similarity	62.5%; Pred. No. 72;
Matches 5; Conservative	0; Mismatches 3; Indels 0; Gaps 0;
OY 1	wrxaxfxg 8
Db 595	WRGRAFTG 602
RESULT 10	H75258
probable 3-demethylubiquinone-9-3-methyltransferase - Deinococcus radiodurans (strain C:Species: Deinococcus radiodurans	
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000	
C:Accession: H75258	
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.V	
M.; Shen, M.; Yamathavan, J.V.; Lam, P.; McDonald, L.; Uteckpack, T.; Zalewski, C.;	
S.; Smith, H.O.; Ventler, J.C.; Fraser, C.M.	
Science 286, 1571-1577, 1999	
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.	
A:Reference number: A75250; MUID:20036896	
A:Accession: H75258	
A:Status: preliminary	
A:Molecule type: DNA	
A:Residues: 1-206 <WHI>	
A:Cross-references: GB:AE002085; GB:AE000513; NID:g6460383; PIDN:AAF12100.1; PID:g646	
A:Experimental source: strain R1	
C:Genetics:	
A:Gene: DR2562	
A:Map position: 1	
Query Match	80.0%; Score 28; DB 2; Length 206;
Best Local Similarity	50.0%; Pred. No. 28;
Matches 4; Conservative	1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
||:||
Db 111 WRTASFDG 118

RESULT 11
A:Accession: A56152
major 2sk outer membrane protein precursor - Brucella abortus
C:Species: Brucella abortus
C:Date: 03-Oct-1995 #sequence_rev1sion 03-Oct-1995 #text_change 08-Oct-1999
C:Accession: A56152
R:De Weiglrose, P.; Lintermans, P.; Limet, J.N.; Cloeckaert, A.
J. Bacteriol. 177, 1911-1914, 1995
A:Title: Cloning and nucleotide sequence of the gene coding for the major 25-kilodalton
A:Reference number: A56152; MUID:95204367
A:Accession: A56152
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-213 <DEA>
A:Cross-references: GB:X79284; NID:g769744; PIDN:CAA55872.1; PID:g769745

Query Match 80.0%; Score 28; DB 2; Length 213;
Best Local Similarity 50.0%; Pred. No. 29;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
||:||
Db 74 WKAGAFAG 81

RESULT 12
T40965
alcohol dehydrogenase (EC 1.1.1.1) class III [similarity] - fission yeast (Schizosacchar
N:Contains: formaldehyde dehydrogenase (glutathione) (EC 1.2.1.1)

C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_rev1sion 03-Dec-1999 #text_change 08-Dec-2000
C:Accession: T40965; T11707
R:Purnelle, B.; Goffeau, A.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, October 1998
A:Reference number: Z21960
A:Accession: T40965
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-380 <PUR>
A:Cross-references: PIDN:CAA21785.1; GSPDB:GN00068; SPDB:SPCC13B11.04c
A:Experimental source: strain 972h-; cosmid c13B11
R:Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, August 1998
A:Reference number: Z17318
A:Accession: T11707
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-61 <SEE>
A:Cross-references: EMBL:AL031532
A:Experimental source: strain 972h(-)
C:Genetics:
A:Gene: SPDB:SPCC13B11.04c
A:Map position: 3; IIR
A:Introns: 9/3
A>Note: SPCC77.01c
A>Note: intron positions not resolved (Incomplete sequence)
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: NAD; oxidoreductase; zinc
F:50.72,179/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

Query Match 80.0%; Score 28; DB 2; Length 380;
Best Local Similarity 62.5%; Pred. No. 51;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 wrxxafxg 8
||:||

Db 319 WRGAFAG 326

RESULT 13
T02995
unspecific monooxygenase (EC 1.14.14.1) - common tobacco
N:Alternate names: cytochrome P450 homolog TBP
C:Species: Nicotiana tabacum (common tobacco)
C:Date: 24-Mar-1999 #sequence_rev1sion 24-Mar-1999 #text_change 21-Jul-2000
C:Accession: T02995
R:Sugita, M.; Sakaki, T.; Yabusaki, Y.; Ohkawa, H.
Biochim. Biophys. Acta 1308, 231-240, 1996
A:Title: Cloning and expression in Escherichia coli and Saccharomyces cerevisiae of a
A:Reference number: Z14816; MUID:96404975
A:Accession: T02995
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-530 <SUG>
A:Cross-references: EMBL:D64052; NID:g1545804; PIDN:BAA10929.1; PID:g1545805
A:Experimental source: strain Bright Yellow 2
C:Genetics:
A:Gene: CTBP
C:Keywords: monooxygenase; oxidoreductase

Query Match 80.0%; Score 28; DB 2; Length 530;
Best Local Similarity 62.5%; Pred. No. 71;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
||:||
Db 254 WRDFAFG 261

RESULT 14
G85842
probable tail component of prophage CP-933V Z3314 [imported] - Escherichia coli (stra
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_rev1sion 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: G85842
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
hiller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: G85842
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-224 <STO>
A:Cross-references: GB:AE005174; NID:g12516364; PIDN:AAG57203.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z3314
C:Superfamily: phage lambda tail assembly protein K

Query Match 77.1%; Score 27; DB 2; Length 224;
Best Local Similarity 50.0%; Pred. No. 53;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
||:||
Db 205 WHASAFVG 212

RESULT 15
E85718
probable tail component of prophage CP-9330 Z2143 [imported] - Escherichia coli (stra
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_rev1sion 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: E85718
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
hiller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda

```
Nature 407, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85718
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-293 <STO>
A:Cross-references: GB:AE005174; NID:gl2515097; PIDN:AAG56209.1; GSPDB:GN00145; UWGP:Z21
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z2143

Query Match      77.1%; Score 27; DB 2; Length 293;
Best Local Similarity 50.0%; Pred. No. 68;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
    | |||
Db 274 WHASFTG 281

RESULT 16
A85817
hypothetical protein Z3081 [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: A85817
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: A85817
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-293 <STO>
A:Cross-references: GB:AE005174; NID:gl2516098; PIDN:AAG56997.1; GSPDB:GN00145; UWGP:Z30
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z3081

Query Match      77.1%; Score 27; DB 2; Length 293;
Best Local Similarity 50.0%; Pred. No. 68;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
    | |||
Db 274 WHASFTG 281

RESULT 17
D64763
alcohol dehydrogenase (EC 1.1.1.1) C - Escherichia coli
N:Alternate names: class III alcohol dehydrogenase
N:Contains: formaldehyde dehydrogenase (glutathione) (EC 1.2.1.1)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 11-Jun-1999
C:Accession: D64763; A42015; S78608
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: D64763
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-369 <BLAT>
A:Cross-references: GB:AE000142; GB:U00096; NID:gl1786542; PIDN:AAC73459.1; PID:gl1786552;
A:Experimental source: strain K-12, substrain MG1655
R:Guthrie, W.G.; Holmquist, B.; Vallee, B.L.
Biochemistry 31, 475-481, 1992
```

```
A:Title: Purification, characterization, and partial sequence of the glutathione-de
A:Reference number: A42015; MUID:92118844
A:Accession: A42015
A:Molecule type: protein
A:Residues: 1-24,'X',26-40,'B',42-45,'G',47 <GUT>
A>Note: this enzyme also has hemithiolacetal dehydrogenase activity
R:Nashimoto, H.; Salto, N.
submitted to the EMBL Data Library, May 1996
A:Description: Kohara library:8F10.
A:Reference number: S78608
A:Accession: S78608
A:Molecule type: DNA
A:Residues: 'MLPLHL',11,'VNRWKSLLTLUHR',26-369 <NAS>
A:Cross-references: EMBL:D85613
A:Experimental source: strain K12
C:Comment: Class III alcohol dehydrogenases are pyrazole-insensitive, are not very
C:Genetics:
A:Gene: adhC
C:Complex: homodimer
C:Function: <ADH>
A:Description: catalyzes oxidation of primary and secondary alcohols to aldehydes
C:Function: <FDH>
A:Description: catalyzes oxidation by NAD+ of formaldehyde and glutathione to S-for
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: alcohol metabolism; homodimer; metalloprotein; NAD; oxidoreductase; z
F:25-360/Domain: long-chain alcohol dehydrogenase homology <LADH>
F:189-218/Region: beta-alpha-beta NAD nucleotide-binding fold
F:40.62.169/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
F:92.95.98.106/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match      77.1%; Score 27; DB 1; Length 369;
Best Local Similarity 50.0%; Pred. No. 85;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
    | |||
Db 309 WRGSAPFG 316

RESULT 18
S57525
alcohol dehydrogenase (EC 1.1.1.1) C - Escherichia coli (isolate VU 3685)
N:Alternate names: class III alcohol dehydrogenase
N:Contains: formaldehyde dehydrogenase (glutathione) (EC 1.2.1.1)
C:Species: Escherichia coli
C:Date: 10-Jul-1992 #sequence_revision 01-May-1998 #text_change 11-Jun-1999
C:Accession: S57525
R:Kuemmerle, N.; Feucht, H.; Kaulfers, P.M.
submitted to the EMBL Data Library, June 1993
A:Description: Plasmid-mediated formaldehyde-resistance in E. coli: nucleotide seq
A:Reference number: S57525
A:Accession: S57525
A:Molecule type: DNA
A:Residues: 1-369 <RUE>
A:Cross-references: EMBL:X73835; NID:g887430; PIDN:CAA52057.1; PID:g887431
A:Experimental source: clinical isolate
C:Comment: Class III alcohol dehydrogenases are pyrazole-insensitive, are not very
C:Complex: homodimer
C:Function: <ADH>
A:Description: catalyzes the oxidation of primary and secondary alcohols to aldehy
C:Function: <FDH>
A:Description: catalyzes the oxidation by NAD+ of formaldehyde and glutathione to
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: alcohol metabolism; homodimer; metalloprotein; NAD; oxidoreductase; z
F:25-360/Domain: long-chain alcohol dehydrogenase homology <LADH>
F:189-218/Region: beta-alpha-beta NAD nucleotide-binding fold
F:40.62.169/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
F:92.95.98.106/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match      77.1%; Score 27; DB 1; Length 369;
Best Local Similarity 50.0%; Pred. No. 85;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 313 WKGTAFGG 320

Qy 1 wrxxafxg 8

I: |||

Db 309 WKGSAFGG 316

Search completed: January 14, 2002, 07:37:32
Job time: 902 sec

RESULT 19

G85530
hypothetical protein adhC [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: G85530
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: G85530
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-369 <STO>
A:Cross-references: GB:AE005174; NID:g12513195; PIDN:AAG54707.1; GSPDB:GN00145; UWGP:Z04
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: adhC
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 77.1%; Score 27; DB 2; Length 369;

Best Local Similarity 50.0%; Pred. No. 85;

Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 wrxxafxg 8

I: |||

Db 309 WKGSAFGG 316

RESULT 20

A33419
alcohol dehydrogenase (EC 1.1.1.1) class III - horse
N:Contains: formaldehyde dehydrogenase (glutathione) (EC 1.2.1.1)
C:Species: Equus caballus (domestic horse)
C:Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 01-May-1998
C:Accession: A33419
R:Kaiser, R.; Holmquist, B.; Vallee, B.L.; Joernvall, H.
Biochemistry 28, 8432-8438, 1989
A:Title: Characteristics of mammalian class III alcohol dehydrogenases, an enzyme less v
A:Reference number: A33419; MUID:90105360
A:Accession: A33419
A:Molecule type: protein
A:Residues: 1-373 <KA>
C:Comment: Class III alcohol dehydrogenases are pyrazole-insensitive, are not very activ
C:Complex: homodimer; does not form heterodimers with the class I alcohol dehydrogenases
C:Function: <ADH>
A:Description: catalyzes the oxidation of primary and secondary alcohols to aldehydes an
C:Function: <EDH>
A:Description: catalyzes the oxidation by NAD+ of formaldehyde and glutathione to S-form
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: acetylated amino end; alcohol metabolism; homodimer; metalloprotein; NAD; ox
F:1-373/Product: alcohol dehydrogenase chi chain #status experimental <NAT>
F:29-364/Domain: long-chain alcohol dehydrogenase homology <LADH>
F:193-222/Region: beta-alpha-beta NAD nucleotide-binding fold
F:1/Modified site: acetylated amino end (Ser) #status experimental
F:44,66,173/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted
F:96,99,102,110/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match 77.1%; Score 27; DB 1; Length 373;

Best Local Similarity 50.0%; Pred. No. 86;

Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 wrxxafxg 8

AC 019005;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-4 (CLOSTRIDIUM PERRINGENS ENTEROTOXIN RECEPTOR) (CPE-
RECEPTOR) (CPE-R).
GN CLDN4 OR CPEPR1 OR CPER
OS Cercopithecus aethiops (Green monkey) (Grivet).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxID=9534;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97242441; PubMed=9087440;
RA Katsuhira J., Inoue N., Horiguchi Y., Matsuda M., Sugimoto N.;
RT "Molecular cloning and functional characterization of the receptor for
Clostridium perfringens enterotoxin";
RL J. Cell Biol. 136:1239-1247(1997).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D88492; BAA22781.1; -;
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin; 1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN; 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 10 30 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 118 138 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
SQ SEQUENCE 209 AA; 22029 MW; 474DB309F95289E CRC64;

Query Match 85.7%; Score 30; DB 1; Length 209;
Best Local Similarity 62.5%; Pred. No. 5.9;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
DB 30 WRVTAFIG 37

RESULT 3
CLD4_HUMAN STANDARD: PRT: 209 AA.
AC 014493;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-4 (CLOSTRIDIUM PERRINGENS ENTEROTOXIN RECEPTOR) (CPE-
RECEPTOR) (CPE-R).
GN CLDN4 OR CPEPR1 OR CPER
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97476271; PubMed=9334247;
RA Sugimoto N., Inoue N., Horiguchi Y., Matsuda M.,

RA Sugimoto N.;
RT "Clostridium perfringens enterotoxin utilizes two structurally related
RT membrane proteins as functional receptors in vivo.";
RL J. Biol. Chem. 272:26552-26558(1997).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB000712; BAA22984.1; -;
DR MIM: 602909; -;
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin; 1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN; 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 118 138 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
SQ SEQUENCE 209 AA; 22077 MW; 0659A93A5F0E4C5 CRC64;

Query Match 85.7%; Score 30; DB 1; Length 209;
Best Local Similarity 62.5%; Pred. No. 5.9;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
DB 30 WRVTAFIG 37

RESULT 4
CLD2_BRARE STANDARD: PRT: 209 AA.
AC Q9YH90;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-LIKE PROTEIN ZP-A9.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Keen T.J., Ingolearn C.F.;
RT Submitted (OCT-1998) to the EMBL/GenBank/DDNJ databases.
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ011790; CAA09778.1; -;
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin; 1.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:40:40 : Search time 37.71 Seconds
(without alignments)
7.778 Million cell updates/sec

Title: 09-185908-1h
Sequence: 1 wrxxafxg 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 50 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	91.4	199	1	VTAK_LAMB
2	30	85.7	209	1	CLD4_CERAE
3	30	85.7	209	1	CLD4_HUMAN
4	30	85.7	209	1	CLD2_BRARE
5	30	85.7	210	1	CLD4_MOUSE
6	30	85.7	219	1	CLD3_MOUSE
7	30	85.7	219	1	CLD3_RAT
8	30	85.7	220	1	CLD3_HUMAN
9	30	85.7	224	1	CLD4_HUMAN
10	29	82.9	367	1	GCST_MYCTU
11	29	82.9	201	1	CS32_ECOLI
12	28	80.0	201	1	OM25_BROV
13	28	80.0	213	1	OM25_BROB
14	28	80.0	213	1	OM25_BRUCA
15	28	80.0	213	1	OM25_BRUNE
16	28	80.0	213	1	OM25_BRUNE
17	28	80.0	213	1	OM25_BRUNE
18	28	80.0	375	1	ADHH_GADMO
19	28	80.0	379	1	FADH_PICPA
20	28	80.0	380	1	FAH2_SCHPO
21	28	80.0	452	1	SPS2_MOUSE
22	27	77.1	208	1	CLD2_BRARE
23	27	77.1	217	1	CLD9_HUMAN
24	27	77.1	217	1	CLD9_MOUSE
25	27	77.1	219	1	CLD6_MOUSE
26	27	77.1	220	1	CLD6_HUMAN
27	27	77.1	369	1	ADH3_ECOLI
28	27	77.1	369	1	ADH3_PASPI
29	27	77.1	373	1	ADHX_HORSE
30	27	77.1	373	1	ADHX_HUMAN
31	27	77.1	373	1	ADHX_MOUSE
32	27	77.1	373	1	ADHX_RABIT
33	27	77.1	373	1	ADHX_RAT

34	27	77.1	373	1	ADHX_UROHA	P80467	uromaslyx h
35	27	77.1	375	1	ADHL_GADMO	P81601	gacus morhu
36	27	77.1	375	1	FADH_PARDE	P45382	paracoccus
37	27	77.1	376	1	ADHI_RHOSH	P72324	rhodobacter
38	27	77.1	376	1	ADHX_MYXGL	P80360	myxine glut
39	27	77.1	376	1	ADHX_SPARA	P79896	sparus auro
40	27	77.1	378	1	ADH3_HAEIN	P44557	haemophilus
41	27	77.1	378	1	ADHX_DROME	P46415	drosophila
42	27	77.1	378	1	ADHX_OCTUV	P81431	octopus vul
43	27	77.1	379	1	ADHX_PEA	P80572	pisum sativ
44	27	77.1	379	1	ADHX_ARATH	P93533	arabidopsis
45	27	77.1	381	1	ADHX_MAIZE	P93629	zea mays (m
46	27	77.1	381	1	ADHX_ORYSA	P93436	oryza sativ
47	27	77.1	381	1	FADH_CANMA	O06099	candida mal
48	27	77.1	384	1	ADHX_CAEEL	Q17335	caenorhabdi
49	27	77.1	386	1	FADH_YEAST	P32771	saccharomyc
50	27	77.1	396	1	STAD_LINUS	P32062	linum usita

ALIGNMENTS

RESULT 1	
ID VTAK_LAMB	STANDARD; PRT; 199 AA.
AC P03729;	
DT 21-JUL-1986 (Rel. 01, Created)	
DT 21-JUL-1986 (Rel. 01, Last sequence update)	
DT 01-APR-1993 (Rel. 25, Last annotation update)	
DE TAIL ASSEMBLY PROTEIN K.	
GN K.	
OS Bacteriophage lambda.	
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;	
CC Lambda phage group.	
OX NCBI_TaxID-10710;	
RN [1]	
RP SEQUENCE FROM N.A.	
RA MEDLINE-83189071; PubMed-6221115;	
RS Sanger F., Coulson A.R., Hong G.F., Hill D.F., Petersen G.B.;	
RT "Nucleotide sequence of bacteriophage lambda DNA.";	
RL J. Mol. Biol. 162:729-773(1982)."	
CC - FUNCTION: GENE K PROTEIN IS INVOLVED IN THE ASSEMBLY OF THE	
CC INITIATOR COMPLEX FOR TAIL POLYMERIZATION. IT HAS NOT BEEN FOUND	
CC IN THE MATURE PHAGE.	
CC - CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-11 IS THE INITIATOR.	
CC This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC the European Bioinformatics Institute. There are no restrictions on its	
CC use by non-profit institutions as long as its content is in no way	
CC modified and this statement is not removed. Usage by and for commercial	
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC or send an email to license@isb-sib.ch).	
DR EMBL; J02459; AAA96551.1; -	
DR PIR; A04355; TUBPKL.	
DR InterPro: IPR000064; NLRP_P60.	
DR Pfam: PF00877; NLRP_P60; I.	
SQ SEQUENCE 199 AA; 23011 MW; CEEB88F01E31ABAE CRC64;	

Query Match 91.4%; Score 32; DB 1; Length 199;
Best Local Similarity 62.5%; Pred. No. 2;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
Db 180 WRSAFTG 187
RESULT 2
CLD4_CERAE STANDARD; PRT; 209 AA.

DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN: 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 81 101 POTENTIAL.
FT TRANSMEM 114 134 POTENTIAL.
FT TRANSMEM 159 179 POTENTIAL.
SQ SEQUENCE 209 AA: 22091 MW: 4413143811853058 CRC64;

Query Match 85.7%; Score 30; DB 1; Length 209;
Best Local Similarity 62.5%; Pred. No. 5.9;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
11111
29 WRVTAFIG 36

RESULT 5
CLD4_MOUSE STANDARD; PRT: 210 AA.
ID CLD4_MOUSE
AC O35054;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE CLAUDIN-4 (CLOSTRIDIUM PERRINGENS ENTEROTOXIN RECEPTOR) (CPE-RECEPTOR) (CPE-R).
GN CLDN4 OR CPETR1 OR CPER.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID:10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:9747621; PubMed:9334247;
RA Katsuhita J., Sugiyama H., Inoue N., Horiguchi Y., Matsuda M., Sugimoto N.;
RT "Clostridium perfringens enterotoxin utilizes two structurally related membrane proteins as functional receptors in vivo."
RL J. Biol. Chem. 272:26652-26658(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE:99110921; PubMed:9892664;
RA Morita K., Furuse M., Fujimoto K., Tsukita S.;
RT "Claudin multigene family encoding four-transmembrane domain protein components of tight junction strands."
RL Proc. Natl. Acad. Sci. U.S.A. 96:511-516(1999).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC
DR EMBL: AB00713; BAA22985.1; -;
DR EMBL: AF087822; AAD09757.1; -;
DR MGD: MGI:1313314; Clnd4.
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF008822; PMP22_Claudin; 1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN: 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.

SQ SEQUENCE 210 AA: 22338 MW: 386D571EC71D6564 CRC64;

Query Match 85.7%; Score 30; DB 1; Length 210;
Best Local Similarity 62.5%; Pred. No. 5.9;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
11111
30 WRVTAFIG 37

RESULT 6
CLD3_MOUSE STANDARD; PRT: 219 AA.
ID CLD3_MOUSE
AC Q920G9;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE CLAUDIN-3 (CLOSTRIDIUM PERRINGENS ENTEROTOXIN RECEPTOR 2) (CPE-RECEPTOR 2) (CPE-R 2).
GN CLDN3 OR CPETR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID:10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:99097345; PubMed:9878248;
RA Paperna T., Peoples R., Wang Y.K., Kaplan P., Francke U.;
RT "Genes for the CPE receptor (CPETR1) and the human homolog of RVPL (CPETR2) are localized within the Williams-Beuren syndrome deletion."
RL Genomics 54:453-459(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE:99110921; PubMed:9892664;
RA Morita K., Furuse M., Fujimoto K., Tsukita S.;
RT "Claudin multigene family encoding four-transmembrane domain protein components of tight junction strands."
RL Proc. Natl. Acad. Sci. U.S.A. 96:511-516(1999).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC
DR EMBL: AF095905; AAD14608.1; -;
DR EMBL: AF087821; AAD09756.1; -;
DR MGD: MGI:1329044; Clnd3.
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF008822; PMP22_Claudin; 1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN: 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 9 29 POTENTIAL.
FT TRANSMEM 81 101 POTENTIAL.
FT TRANSMEM 116 136 POTENTIAL.
FT TRANSMEM 160 180 POTENTIAL.
SQ SEQUENCE 219 AA: 23284 MW: 62F67810D9B9BD37 CRC64;

Query Match 85.7%; Score 30; DB 1; Length 219;
Best Local Similarity 62.5%; Pred. No. 6.2;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxafxg 8

```
Db 29 WRVSAPFG 36

RESULT 7
CLD3_RAT STANDARD: PRT: 219 AA.
ID CLD3_RAT
AC Q63400;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE CLAUDIN-3 (VENTRAL PROSTATE.1 PROTEIN) (RVP1).
GN CLDN3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92130987; PubMed=1723140;
RA Brieht M.M., Miesfeld R.L.;
RT "Isolation and characterization of transcripts induced by androgen
withdrawal and apoptotic cell death in the rat ventral prostate.";
RL Mol. Endocrinol. 5:1381-1388(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX Keen T.J., Ingelhearn C.F.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M74067; AAA41760.1; -
DR EMBL: AJ011656; CAA09727.1; -
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR00729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN.
KW Tight junction; Transmembrane.
FT TRANSMEM 9 29
FT TRANSMEM 81 101 POTENTIAL.
FT TRANSMEM 116 136 POTENTIAL.
FT TRANSMEM 162 182 POTENTIAL.
FT CONFLICT 4 4 G -> S (IN REF. 1).
FT CONFLICT 55 55 MISSING (IN REF. 1).
FT CONFLICT 217 219 DTV -> TTSERGARPHHHYQPSWYTPRPACLSAETTP
PPSRRLQTPRSLARLEEDRQGVFSPVAT (IN REF.
1).
SQ SEQUENCE 219 AA; 23314 MW; 820CC6BFC20D122D CRC64;

Query Match 85.7%; Score 30; DB 1; Length 219;
Best Local Similarity 62.5%; Pred. No. 6.2;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 WRXSAPFG 8
ID CLD3_HUMAN STANDARD: PRT: 220 AA.
AC O15551.*
```

```
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-3 (CLOSTRIDIUM PERFRINGENS ENTEROTOXIN RECEPTOR 2) (CPE-
RECEPTOR 2) (CPE-R 2) (VENTRAL PROSTATE.1 PROTEIN HOMOLOG) (HRVP1).
GN CLDN3 OR CPERR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98110580; PubMed=9441748;
RA Peacock R.E., Keen T.J., Ingelhearn C.F.;
RT "Analysis of a human gene homologous to rat ventral prostate.1
protein.";
RL Genomics 46:443-449(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97476271; PubMed=9334247;
RA Katsuhira J., Sugiyama H., Inoue N., Horiguchi Y., Matsuda M.,
Sugimoto N.;
RT "Clostridium perfringens enterotoxin utilizes two structurally related
membrane proteins as functional receptors in vivo.";
RL J. Biol. Chem. 272:26652-26658(1997).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF007189; AAC78277.1; -
DR EMBL: AB000714; BAA22986.1; -
DR MTM: 602910; -
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR00729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN.
KW Tight junction; Transmembrane.
FT TRANSMEM 9 29
FT TRANSMEM 81 101 POTENTIAL.
FT TRANSMEM 116 136 POTENTIAL.
FT TRANSMEM 160 180 POTENTIAL.
SQ SEQUENCE 220 AA; 23318 MW; 1C826EFFE1563C56 CRC64;

Query Match 85.7%; Score 30; DB 1; Length 220;
Best Local Similarity 62.5%; Pred. No. 6.2;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 WRXSAPFG 8
ID CLD3_HUMAN STANDARD: PRT: 224 AA.
AC P56750;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE CLAUDIN-17.
GN CLDN17.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
```

```

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RA SEQUENCE FROM N.A.
RA Keen T.J., Inglehearn C.F.;
RA Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; Pubmed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Seda E.,
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald A., Rump J., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudon J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyma S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordstam G., Horstischer K., Brandt P.,
RA Scharte M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Resselmann L., Dagand E.,
RA Wehmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehnach H., Reinhardt R., Yaspo M.-L.;
RA "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: A1250712; CAB60616.1; -.
DR EMBL: AF001707; BAA95566.1; -.
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin; 1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN; 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 125 145 POTENTIAL.
FT TRANSMEM 165 185 POTENTIAL.
SQ SEQUENCE 224 AA; 24603 MW; 1833ED3178B7F63A CRC64;

```

```

Query Match 85.7% Score 30; DB 1; Length 224;
Best Local Similarity 62.5%; Pred. No. 6.3;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

OY 1 wrxxafxg 8
   || || |
Db 30 WRVSAFVG 37

```

```

RESULT 10
GCST_MYCTU STANDARD; PRT: 367 AA.
AC Q10376;
DT 01-OCT-1996 (Rel. 34, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE AMINOMETHYLTRANSFERASE (EC 2.1.2.10) (GLYCINE CLEAVAGE SYSTEM
DE T PROTEIN)
DE GCVT OR RV2211C OR MT2267 OR MTCT190.22.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriales; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; Pubmed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Ruster S., Seeger K., Skellon S., Squares S., Squares R.,
RA Sutton J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Hatt D., Hickey E.,
RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE GLYCINE CLEAVAGE SYSTEM CATALYSES THE DEGRADATION OF
CC GLYCINE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: (6S)-TETRAHYDROFOLATE +
CC S-AMINOMETHYLHYDROLIPOYLPROTEIN -> (6R)-5,10-
CC METHYLENETHYLDIHYDROLIPOYL + NH(3) + DIHYDROLIPOYLPROTEIN.
CC -1- SUBUNIT: THE GLYCINE CLEAVAGE SYSTEM IS COMPOSED OF FOUR PROTEINS:
CC P, T, L, AND H (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE GCVT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Z70283; CA94254.1; ALT_INIT.
DR EMBL: AE007072; AAK46553.1; ALT_INIT.
DR TIGR: MT2267; -.
DR Tuberculist; RV2211c; -.
DR InterPro: IPR002536; GCVT.
DR Pfam: PF01571; GCVT; 1.
KW transferase; Aminotransferase; Complete proteome.
SQ SEQUENCE 367 AA; 38317 MW; 47CD3E1FFE46B321 CRC64;

```

```

Query Match 82.9% Score 29; DB 1; Length 367;
Best Local Similarity 62.5%; Pred. No. 17;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

OY 1 wrxxafxg 8
   || || |
Db 263 WRKDAFFG 270

```

```

RESULT 11
CS32_ECOLI STANDARD; PRT: 937 AA.
ID CS32_ECOLI
AC P15484; P15485; P15486; P15487;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE OUTER MEMBRANE USHER PROTEIN CS3-2 PRECURSOR (CS3 PILI SYNTHESIS 104
DE KDA PROTEIN).
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

```

```

OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P3176;
RX MEDLINE=90158116; PubMed=2576094;
RA JalaJakumari M.B., Thomas C.J., Halter R., Manning P.A.;
RT "Genes for biosynthesis and assembly of CS3 pili of CFA/II
RT enterotoxigenic Escherichia coli: novel regulation of pilus
RT production by bypassing an amber codon.";
RL Mol. Microbiol. 3:1685-1695(1989).
CC -1- FUNCTION: THIS PROTEIN IS ESSENTIAL FOR THE BIOGENESIS OF
CC MATURE CS3 PILI.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE
CC (BY SIMILARITY)
CC -1- ALTERNATIVE PRODUCTS: FIVE PROTEIN ARE PRODUCED BY ALTERNATIVE
CC INITIATION. THE FIFTH PROTEIN REQUIRES THE SUPPRESSION OR
CC READTHROUGH OF AN INTERNAL AMBER CODON IN POSITION 754.
CC -1- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X16944; CAA34817.1; -.
DR EMBL: X16944; CAA34816.1; ALT_INIT.
DR InterPro: IPR000015; Fimb_usher.
DR Pfam: PF00577; Usher: 1.
DR PROSITE: PS01151; FIMBRIAL_USHER: 1.
KW Outer membrane; Transmembrane; Fimbrin; Transport; Signal;
KW Alternative initiation.
FT SIGNAL 1 ?
FT CHAIN 1 937 CS3 PILI SYNTHESIS 104 KDA PROTEIN.
FT CHAIN 181 753 CS3 PILI SYNTHESIS 63 KDA PROTEIN.
FT CHAIN 317 753 CS3 PILI SYNTHESIS 48 KDA PROTEIN.
FT CHAIN 451 753 CS3 PILI SYNTHESIS 33 KDA PROTEIN.
FT CHAIN 572 753 CS3 PILI SYNTHESIS 20 KDA PROTEIN.
SQ SEQUENCE 937 AA; 104150 MW; 6C53056EE96277B5 CRC64;

Query Match 82.9%; Score 29; DB 1; Length 937;
Best Local Similarity 62.5%; Pred. No. 40;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
| | | |
Db 595 WRCGAFIG 602

RESULT 12
OM25_BRUV STANDARD; PRT; 201 AA.
ID OM25_BRUV
AC 045335;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 25 KDA OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
GN OMP25.
OS Brucella ovis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bruciellaceae; Brucella.
OX NCBI_TaxID=236;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=63/290;
RX MEDLINE=96239016; PubMed=8675306;
RA Cloeckaert A., Verger J.M., Grayson M., Zygmunt M.S., Grepinet O.;
RT "Nucleotide sequence and expression of the gene encoding the major
RT 25-kilodalton outer membrane protein of Brucella ovis: Evidence for

```

```

RT antigenic shift, compared with other Brucella species, due to a
RT deletion in the gene.";
RL Infect. Immun. 64:2047-2055(1996).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY. HAS A C-TERMINAL
CC DELETION COMPARED TO THAT OF OTHER BRUCELLA SPECIES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U33004; AAB06702.1; -.
DR InterPro: IPR000498; Ompa_tmemb.
DR Pfam: PF01389; Ompa_membrane; 1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 23
FT CHAIN 24 201 POTENTIAL.
SQ SEQUENCE 201 AA; 21817 MW; 85A4897489A0935B CRC64;

Query Match 80.0%; Score 28; DB 1; Length 201;
Best Local Similarity 50.0%; Pred. No. 16;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
| | | |
Db 74 WKAGAFAG 81

RESULT 13
OM25_BRUV STANDARD; PRT; 213 AA.
ID OM25_BRUV
AC 044664;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 25 KDA OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
GN OMP25.
OS Brucella abortus.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bruciellaceae; Brucella.
OX NCBI_TaxID=235;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=544S;
RX MEDLINE=95204367; PubMed=7896724;
RA de Wergifosse P., Lintermans P., Linet J.N., Cloeckaert A.;
RT "Cloning and nucleotide sequence of the gene coding for the major 25-
RT kilodalton outer membrane protein of Brucella abortus.";
RL J. Bacteriol. 177:1911-1914(1995).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X79284; CAA55872.1; -.
DR InterPro: IPR000498; Ompa_tmemb.
DR Pfam: PF01389; Ompa_membrane; 1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 23
FT CHAIN 24 213 POTENTIAL.
SQ SEQUENCE 25 KDA OUTER-MEMBRANE IMMUNOGENIC

```



```
SO SEQUENCE 213 AA: 23052 MW: 2328515F1F794BC7 CRC64:
Query Match
Best Local Similarity 80.0%; Score 28; DB 1; Length 213;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
   1: 111
   74 WKAGAFAG 81

DB 74 WKAGAFAG 81

RESULT 14
OM25_BRUC  STANDARD: PRT; 213 AA.
ID OM25_BRUC  STANDARD: PRT; 213 AA.
AC Q45110;
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 25 KDA OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
OS OMP25.
SN Brucella canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
CC Brucellaceae; Brucella.
CX NCBI_TaxID:36855;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RM6/66;
RX MEDLINE:96239016; PubMed=8675306;
RA Cloeckaert A., Verger J.M., Grayon M., Zygmunt M.S., Grepinet O.;
RT "Nucleotide sequence and expression of the gene encoding the major
RT 25-kilodalton outer membrane protein of Brucella ovis: Evidence for
RT antigenic shift, compared with other Brucella species, due to a
RT deletion in the gene."
RL Infect. Immun. 64:2047-2055(1996).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: U39358; AAB36692.1; -.
DR InterPro: IPR000498; OmpA_tmem.
DR Pfam: PF01389; OmpA_membrane.1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 213 25 KDA OUTER-MEMBRANE IMMUNOGENIC
FT PROTEIN.
SQ SEQUENCE 213 AA: 23123 MW: D7E40E247A39B9DF CRC64:

Query Match
Best Local Similarity 80.0%; Score 28; DB 1; Length 213;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
   1: 111
   74 WKAGAFAG 81

DB 74 WKAGAFAG 81

RESULT 15
OM25_BRUC  STANDARD: PRT; 213 AA.
ID OM25_BRUC  STANDARD: PRT; 213 AA.
AC Q45321;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 25 KDA OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
OS OMP25.
SN Brucella canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
CC Brucellaceae; Brucella.
CX NCBI_TaxID:36855;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5K33;
RX MEDLINE:96239016; PubMed=8675306;
RA Cloeckaert A., Verger J.M., Grayon M., Zygmunt M.S., Grepinet O.;
RT "Nucleotide sequence and expression of the gene encoding the major
RT 25-kilodalton outer membrane protein of Brucella ovis: Evidence for
RT antigenic shift, compared with other Brucella species, due to a
RT deletion in the gene."
RL Infect. Immun. 64:2047-2055(1996).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
GN OMP25.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
CC Brucellaceae; Brucella.
CX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M;
RX MEDLINE=96239016; PubMed=8675306;
RA Cloeckaert A., Verger J.M., Grayon M., Zygmunt M.S., Grepinet O.;
RT "Nucleotide sequence and expression of the gene encoding the major
RT 25-kilodalton outer membrane protein of Brucella ovis: Evidence for
RT antigenic shift, compared with other Brucella species, due to a
RT deletion in the gene."
RL Infect. Immun. 64:2047-2055(1996).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: U39359; AAB36694.1; -
 DR InterPro: IPR000498; OmpA_tmemb.
 DR Pfam: PF01389; OmpA_membrane; 1.
 KW Antigen; Outer membrane; Signal.
 FT SIGNAL 1 23
 FT CHAIN 24 213
 FT POTENTIAL.
 FT 25 KDA OUTER-MEMBRANE IMMUNOGENIC
 FT PROTEIN.
 SO SEQUENCE 213 AA; 23167 MW; 49EE0F47B784F87 CRC64;

Query Match 80.0%; Score 28; DB 1; Length 213;
 Best Local Similarity 50.0%; Pred. No. 17;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
 | : | | |
 Db 74 WKAGAFAG 81

RESULT 17
 OM25_BRUSU STANDARD; PRT; 213 AA.
 AC 045689;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE 25 KDA OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
 GN OMP25.
 OS Brucella suis.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Brucellaceae; Brucella.
 OX NCBI_TaxID=29461;
 RX [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=1330;
 RC MEDLINE=96239016; PubMed=8675306;
 RA Cleeckaert A., Verger J.M., Grayon M., Zygmunt M.S., Grepinet O.;
 "Nucleotide sequence and expression of the gene encoding the major
 25-kilodalton outer membrane protein of Brucella ovis: Evidence for
 antigenic shift, compared with other Brucella species, due to a
 deletion in the gene.";
 RT Infect. Immun. 64:2047-2055(1996).
 RL -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE OMP25 / ROPE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL: U39397; AAB36695.1; -
 DR InterPro: IPR000498; OmpA_tmemb.
 DR Pfam: PF01389; OmpA_membrane; 1.
 KW Antigen; Outer membrane; Signal.
 FT SIGNAL 1 23
 FT CHAIN 24 213
 FT POTENTIAL.
 FT 25 KDA OUTER-MEMBRANE IMMUNOGENIC
 FT PROTEIN.
 SO SEQUENCE 213 AA; 23151 MW; 55F33CF46F6D0ED3 CRC64;

Query Match 80.0%; Score 28; DB 1; Length 213;
 Best Local Similarity 50.0%; Pred. No. 17;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
 | : | | |
 Db 74 WKAGAFAG 81

RESULT 18
 ADH1_GADMO STANDARD; PRT; 375 AA.
 ID ADH1_GADMO
 AC P81600;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ALCOHOL DEHYDROGENASE CLASS III H CHAIN (EC 1.1.1.1) (GLUTATHIONE-
 DEPENDENT FORMALDEHYDE DEHYDROGENASE) (EC 1.2.1.1) (FDH).
 OS Gadus morhua (Atlantic cod).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae;
 OC Gadus.
 OX NCBI_TaxID=8049;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=97085413; PubMed=8931553;
 RA Danielsson O., Shafiqat J., Estonius M., El-Ahmad M., Joernvall H.;
 "Isocysteine multiplicity with anomalous dimer patterns in a class III
 RT alcohol dehydrogenase. Effects on the activity and quaternary
 RT structure of residue exchanges at 'non-functional' sites in a native
 RT protein.";
 RL Biochemistry 35:14561-14568(1996).
 CC -1- FUNCTION: CLASS-III ADH IS REMARKABLY INEFFECTIVE IN OXIDIZING
 CC ETHANOL, BUT IT READILY CATALYZES THE OXIDATION OF LONG-CHAIN
 CC PRIMARY ALCOHOLS AND THE OXIDATION OF S-(HYDROXYMETHYL)
 CC GLUTATHIONE.
 CC -1- CATALYTIC ACTIVITY: ALCOHOL + NAD(+) = ALDEHYDE OR KETONE + NADH.
 CC -1- CATALYTIC ACTIVITY: FORMALDEHYDE + GLUTATHIONE + NAD(+) =
 CC S-FORMYLGLUTATHIONE + NADH.
 CC -1- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY.
 CC -1- SUBUNIT: HOMODIMER AND HETERODIMER WITH L CHAIN.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
 CC FAMILY. CLASS-III SUBFAMILY.
 CC InterPro: IPR002328; ADH_Zinc.
 DR InterPro: IPR002085; Adh_Zinc.
 DR Pfam: PF00107; adh_zinc; 1.
 DR PROSITE: PS00059; ADH_ZINC; 1.
 KW Oxidoreductase; Zinc; NAD; Multigene family; Acetylation.
 FT MOD_RES 1
 FT METAL 46 46
 FT METAL 68 68
 FT METAL 98 98
 FT METAL 101 101
 FT METAL 104 104
 FT METAL 112 112
 FT METAL 175 175
 FT BINDING 116 116
 FT IMPORTANT FOR FDH ACTIVITY AND ACTIVATION
 FT BY FATTY ACIDS (BY SIMILARITY).
 SO SEQUENCE 375 AA; 39669 MW; 0B9760AB77329FE3 CRC64;

Query Match 80.0%; Score 28; DB 1; Length 375;
 Best Local Similarity 50.0%; Pred. No. 29;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
 | : | | |
 Db 315 WKATAFAG 322

RESULT 19
 FADH_PICPA STANDARD; PRT; 379 AA.
 ID FADH_PICPA
 AC 074685;
 DT 20-AUG-2001 (Rel. 40, Created)

```

DT 20-Arg-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE GLUTATHIONE-DEPENDENT FORMALDEHYDE DEHYDROGENASE (EC 1.2.1.1) (FDH)
GN (FDH) (FLDH).
OS Pichia pastoris (Yeast).
CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC Saccharomycetales; Saccharomycetaceae; Pichia.
CC NCBI_TaxID=4922;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCI100;
RX MEDLINE=98382577; PubMed=9714758;
RA Shen S., Sutter G., Jeffries T.W., Cregg J.M.;
RT "A strong nitrogen source-regulated promoter for controlled expression
of foreign genes in the yeast Pichia pastoris.";
RL Gene 216:93-102(1998).
CC -1- CATALYTIC ACTIVITY: FORMALDEHYDE + GLUTATHIONE + NAD(+) = S-
FORMYLGLUTATHIONE + NADH.
CC -1- COPACITOR: REQUIRES ZINC FOR ITS ACTIVITY (POTENTIAL)
CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
FAMILY, CLASS-III SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF066054; AAC35913.1; -.
DR HSSP: P11766; ITEX.
DR InterPro: IPR002328; ADH_zinc.
DR InterPro: IPR002085; Adh_zinc.
DR InterPro: IPR00205; NAD_binding.
DR Pfam: PF00107; adh_zinc.1.
DR PROSITE: PS00059; ADH_ZINC; 1.
KW Oxidoreductase; Zinc; NAD.
FT METAL 47 47 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 69 69 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 99 99 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 102 102 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 105 105 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 113 113 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 176 176 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 379 AA; 40559 MW; CEB8E9229A54BDE2 CRC64;

Query Match 80.0%; Score 28; DB 1; Length 379;
Best Local Similarity 62.5%; Pred. No. 29;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
|| || |
Db 318 WRCAFGC 325

RESULT 20
FAH2_SCHPO STANDARD: PRT; 380 AA.
AC 074540;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE GLUTATHIONE-DEPENDENT FORMALDEHYDE DEHYDROGENASE 2
DE (EC 1.2.1.1) (FDH) (FLDH) (FLD).
GN SPC13B11.04C OR SPC177.01C.
OS Schizosaccharomyces pombe (Fission Yeast).
CC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
CC Schizosaccharomycetales; Schizosaccharomycetaceae;
CC NCBI_TaxID=4896;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Purnelle B., Goffeau A., Lyne M., Rajandream M.A., Barrell B.G.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-61 FROM N.A.
RC STRAIN=972;
RA Seeger K., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: FORMALDEHYDE + GLUTATHIONE + NAD(+) =
S-FORMYLGLUTATHIONE + NADH.
CC -1- COPACITOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
FAMILY, CLASS-III SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AL032681; CA21785.1; -.
DR EMBL: AL031532; CA20705.1; -.
DR HSSP: P11766; ITEX.
DR InterPro: IPR002328; ADH_zinc.
DR InterPro: IPR002085; Adh_zinc.
DR Pfam: PF00107; adh_zinc.1.
DR PROSITE: PS00059; ADH_ZINC; 1.
KW Hypothetical protein; Oxidoreductase; Zinc; NAD.
FT METAL 50 50 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 72 72 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 102 102 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 105 105 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 108 108 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 116 116 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 179 179 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 380 AA; 40712 MW; 2E39F0F347BBBC45 CRC64;

```

```

Query Match 80.0%; Score 28; DB 1; Length 380;
Best Local Similarity 62.5%; Pred. No. 29;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
|| || |
Db 319 WRCAFGC 326

```

Search completed: January 14, 2002, 07:40:40
Job time: 506 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:39:49 ; Search time 112.89 Seconds
(without alignments)
10.366 Million cell updates/sec

Title: 09-185908-1h
Perfect score: 35
Sequence: 1 wrxxafg 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database :
1: SP_TREMBL_17:*
2: SP_Archea:*
3: SP_Bacteria:*
4: SP_Fungi:*
5: SP_Human:*
6: SP_Invertebrate:*
7: SP_Mammal:*
8: SP_Mhc:*
9: SP_Organelle:*
10: SP_Phage:*
11: SP_Plant:*
12: SP_Rodent:*
13: SP_Virus:*
14: SP_Vertebrate:*
15: SP_Unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	31	88.6	266	2	Q9EWM0
2	30	85.7	159	9	Q38144
3	30	85.7	214	13	Q9DE12
4	30	85.7	214	13	Q98SR2
5	30	85.7	330	2	Q9S2V4
6	30	85.7	370	2	Q9HY01
7	30	85.7	372	2	Q9RD05
8	29	82.9	183	2	P97158
9	29	82.9	303	2	P97157
10	29	82.9	613	2	Q9A956
11	28	80.0	206	2	Q9RRD0
12	28	80.0	479	12	Q9Q0N6
13	28	80.0	479	12	Q9UH75
14	28	80.0	497	4	Q9BRR6
15	28	80.0	530	10	Q04892
16	27	77.1	59	4	Q9BUQ2
17	27	77.1	64	7	Q30824
18	27	77.1	247	2	Q9EYEA
19	27	77.1	369	2	Q9Y399

20	27	77.1	369	2	Q24687	Q24687 anaebacter
21	27	77.1	369	2	Q9A5D4	Q9A5D4 caulobacter
22	27	77.1	377	5	Q9NUD0	Q9NUD0 branchiosto
23	27	77.1	377	5	Q9NUC3	Q9NUC3 branchiosto
24	27	77.1	377	5	Q9BJ34	Q9BJ34 branchiosto
25	27	77.1	377	5	Q9BJ33	Q9BJ33 branchiosto
26	27	77.1	378	2	Q9JRB0	Q9JRB0 neisseria m
27	27	77.1	379	10	Q9END2	Q9END2 arabidopsis
28	27	77.1	396	10	Q82014	Q82014 linum usita
29	27	77.1	396	10	Q9SBA2	Q9SBA2 linum usita
30	27	77.1	1083	4	Q14868	Q14868 homo sapien
31	27	77.1	1083	11	Q63744	Q63744 rattus norv
32	27	77.1	1091	4	Q43199	Q43199 homo sapien
33	27	77.1	1092	11	Q9R029	Q9R029 mus musculu
34	27	77.1	1534	4	Q9C0E0	Q9C0E0 homo sapien
35	26	74.3	64	7	Q30825	Q30825 ovis aries
36	26	74.3	65	11	Q9Z2L5	Q9Z2L5 mus musculu
37	26	74.3	91	7	Q19588	Q19588 homo sapien
38	26	74.3	149	5	Q9NMU3	Q9NMU3 leishmania
39	26	74.3	154	2	Q50394	Q50394 mycobacteri
40	26	74.3	165	7	Q9MW45	Q9MW45 homo sapien
41	26	74.3	169	1	Q58027	Q58027 pyrococcus
42	26	74.3	171	1	Q9UY30	Q9UY30 pyrococcus
43	26	74.3	200	2	Q9A697	Q9A697 caulobacter
44	26	74.3	212	2	Q49799	Q49799 mycobacteri
45	26	74.3	212	7	Q95733	Q95733 homo sapien
46	26	74.3	213	7	Q19633	Q19633 homo sapien
47	26	74.3	229	9	Q9ZK28	Q9ZK28 mycobacteri
48	26	74.3	232	2	Q53729	Q53729 mycobacteri
49	26	74.3	248	2	Q9RCY8	Q9RCY8 streptomyce
50	26	74.3	260	2	Q9K0B7	Q9K0B7 vibrio chol

ALIGNMENTS

RESULT 1
Q9EWM0 PRELIMINARY: PRT: 266 AA.
AC Q9EWM0;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE POTATIVE ENOYL-COA HYDRATASE.
GN 25CK31.11C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RX [1]
RN SEQUENCE FROM N.A.
RC SRRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RP SRRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.
RN [3]
RN SEQUENCE FROM N.A.
RP SRRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denaplatte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL: AL512667; CAC21620.1; -;
DR InterPro: IPR001753; Enoyl-CoA_hydratase.
DR Pfam: PF00378; ECH: 1.
DR PROSITE: PS00166; Enoyl-CoA_HYDRATASE: 1.
SQ SEQUENCE 266 AA: 28121 MW: C07F9346B82E0451 CRC64:

```

Query Match                      88.6%; Score 31; DB 2; Length 266;
Best Local Similarity 62.5%; Pred. NO. 29;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
   |||||
Db 239 WRSVAFSG 246

RESULT 2
ID Q38144 PRELIMINARY; PRT; 159 AA.
AC Q38144;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE REQUIRED FOR REPLICATION INITIATION.
GN 36.
OS Bacteriophage SPp1.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;
OC Lambda phage group.
OX NCBI_TaxID=10724;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94172631; PubMed=8126723;
RA Pedre X., Weise F., Chai S., Lueder G., Alonso J.C.;
RT "Analysis of cis and trans acting elements required for the initiation
RT of DNA replication in the Bacillus subtilis bacteriophage SPp1.";
RL J. Mol. Biol. 236:1324-1340(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Alonso J.C., Lueder G., Stiege A.C., Chai S., Weise F., Trautner T.A.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Alonso J.C.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: X67865; CAA48055.1; -
DR EMBL: X97918; CAA66491.1; -
DR HSSP: P02339; IEYG.
DR InterPro: IPR000424; SSB.
DR Pfam: PF00436; SSB; 1.
SQ SEQUENCE 159 AA; 17088 MW; C575E3494C37B589 CRC64;

Query Match                      85.7%; Score 30; DB 9; Length 159;
Best Local Similarity 62.5%; Pred. NO. 29;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
   |||||
Db 77 WRTGAFEG 84

RESULT 3
ID Q9DE12 PRELIMINARY; PRT; 214 AA.
AC Q9DE12;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE TRANSMEMBRANE TIGHT JUNCTION PROTEIN CLAUDIN.
GN CLA.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
OC Xenopodidae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RL TISSUE=DORSAL LIP;

```

```

RA Brizuela B.J., Messely O., De Robertis E.M.;
RT "Overexpression of Xenopus Claudin Tight Junction Protein Affects Cell
RT Adhesion and Causes Randomization of the Left-Right Body Axis.";
RL Dev. Biol. 0:0-0(2001).
DR EMBL: AF224712; AAG44257.1; -
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR000729; PMP22_Claudin.
DR Pfam: PF00822; PMP22_Claudin; 1.
DR PRINTS: PR01077; CLAUDIN.
DR PROSITE: PS01346; CLAUDIN; UNKNOWN_1.
SQ SEQUENCE 214 AA; 23151 MW; 3F0D65E3ADE311E6 CRC64;

Query Match                      85.7%; Score 30; DB 13; Length 214;
Best Local Similarity 62.5%; Pred. NO. 38;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
   |||||
Db 30 WRTVAFIG 37

RESULT 4
ID Q98SR2 PRELIMINARY; PRT; 214 AA.
AC Q98SR2;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CLAUDIN-3.
GN CLDN3.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Reardon E., Kojima S., Rizzolo L.J.;
RT "Sequence of chick claudin-3 cDNA.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF34677; AAK20876.1; -
SQ SEQUENCE 214 AA; 23090 MW; A2540116CBD53978 CRC64;

Query Match                      85.7%; Score 30; DB 13; Length 214;
Best Local Similarity 62.5%; Pred. NO. 38;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
   |||||
Db 29 WRTVAFIG 36

RESULT 5
ID Q9S2V4 PRELIMINARY; PRT; 330 AA.
AC Q9S2V4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE PUTATIVE MEMBRANE PROTEIN.
GN SCA410.35C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
RN [2]

```

RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Rodenbach M., Kleser H.M., Denaplatte D., Eichner A., Cullum J.,
 RT Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb *Scripomyces coelicolor* A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL; AL109663; CAB52011.1; -
 SO SEQUENCE 330 AA; 35177 MW; DB21004753E11700 CRC64;

Query Match 85.7%; Score 30; DB 2; Length 330;
 Best Local Similarity 62.5%; Pred. No. 59;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
 DB 31 WRLLAFAG 38

RESULT 6
 O9HY01 PRELIMINARY; PRT; 370 AA.

DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE ALCOHOL DEHYDROGENASE CLASS III.

GN ADHC OR PA3629.

OC Pseudomonas aeruginosa.

OS Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.

NCBI_TaxID=287;

RP SEQUENCE FROM N.A.

RC STRAIN=PA01.

MDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,

Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,

Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,

Brody L.L., Coulter S.N., Folger K.R., Kas A., Larijs K., Lim R.M.,

Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,

Reizer J., Sailer M.H., Hancock R.E.W., Jory S., Olson M.V.;

"Complete genome sequence of *Pseudomonas aeruginosa* PA01, an

opportunistic pathogen.";

Nature 406:959-964(2000).

-1- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).

-1- SIMILARITY: TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE FAMILY.

EMBL: AE004783; AAC07017.1; -

InterPro: IPR002085; Adh_zinc.

DR InterPro: IPR002328; Adh_zinc.

Pfam: PF00107; adh_zinc.1.

PROSITE: PS00059; ADH_ZINC.1.

Complete proteome: Oxidoreductase; Zinc.

SEQUENCE 370 AA; 39208 MW; 5B48DC94FB45DE10 CRC64;

Query Match 85.7%; Score 30; DB 2; Length 370;
 Best Local Similarity 62.5%; Pred. No. 66;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
 DB 310 WRGSAFCG 317

RESULT 7

O9RD05
 ID O9RD05 PRELIMINARY; PRT; 372 AA.
 AC O9RD05;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE PUTATIVE FORMALDEHYDE DEHYDROGENASE (GLUTATHIONE-DEPENDENT).
 GN Fdh.
 OS Pseudomonas sp.
 OC Bacteria; Proteobacteria.
 NCBI_TaxID=306;
 RP SEQUENCE FROM N.A.
 RC STRAIN=HR199;
 RA Priefer H., Overhage J., Steinbuechel A.;
 RT "Identification and molecular characterization of the eugonol
 hydroxylase genes (ehyA/ehyB) of *Pseudomonas* sp. strain HR199.";
 RL Submitted (Jul-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN=HR199;
 RX MEDLINE=97252489; PubMed=9098058;
 RA Priefer H., Rabenhorst J., Steinbuechel A.;
 RT "Molecular characterization of genes of *Pseudomonas* sp. strain HR199
 involved in bioconversion of vanillin to protocatechuate.";
 RT J. Bacteriol. 179:2595-2607(1997).

CC -1- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).

-1- SIMILARITY: TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE FAMILY.

EMBL: AJ243941; CAB64351.1; -

HSSP: P11766; ITEX.

InterPro: IPR002085; Adh_zinc.

DR InterPro: IPR002328; Adh_zinc.

Pfam: PF00107; adh_zinc.1.

PROSITE: PS00059; ADH_ZINC.1.

Oxidoreductase; Zinc.

SEQUENCE 372 AA; 39343 MW; AAF0B44FBE8426E2 CRC64;

Query Match 85.7%; Score 30; DB 2; Length 372;
 Best Local Similarity 62.5%; Pred. No. 67;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
 DB 310 WRGSAFCG 317

RESULT 8
 P97158 PRELIMINARY; PRT; 182 AA.

DT 01-MAY-1997 (TREMBlrel. 03, Created)

DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE 20 KDA PROTEIN.

OS *Escherichia coli*.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC *Escherichia*.

NCBI_TaxID=562;

RP SEQUENCE OF 122-303 FROM N.A.

RC STRAIN=PB176;

MDLINE=90158116; PubMed=2576094;

RA Jaisakumari M.B., Thomas C.J., Halter R., Manning P.A.;

"Genes for biosynthesis and assembly of CS3 pill of GFA/II

enterotoxigenic *Escherichia coli*: novel regulation of pilus production

by bypassing an amber codon.";

Mol. Microbiol. 3:1685-1695(1989).

EMBL: X16944; CAA34819.1; -

InterPro: IPR000015; Pfam_usher.

Pfam: PF00577; Usher.1.

SEQUENCE 182 AA; 19604 MW; B39386E287B03B2 CRC64;

Query Match 82.9%; Score 29; DB 2; Length 182;
Best Local Similarity 62.5%; Pred. No. 54;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxafxg 8
|||
Db 24 WRGRAFIG 31

RESULT 9

ID P97157 PRELIMINARY; PRT; 303 AA.

AC P97157;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 33 KDA PROTEIN.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PB176;
RX MEDLINE=90158116; Pubmed=2576094;
RA JalaJakumari M.B., Thomas C.J., Halter R., Manning P.A.;
RT "Genes for biosynthesis and assembly of CS3 pili of CFA/II
RT enterotoxigenic Escherichia coli: novel regulation of pilus production
RT by bypassing an amber codon.";
RL Mol. Microbiol. 3:1685-1695(1989).
DR InterPro: IPR000015; FimB_usher.
DR Pfam: PF00577; Usher; 1.
SQ SEQUENCE 303 AA; 32839 MW; A8C62A5A60164542 CRC64;

Query Match 82.9%; Score 29; DB 2; Length 303;
Best Local Similarity 62.5%; Pred. No. 90;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxafxg 8
|||
Db 145 WRGRAFIG 152

RESULT 10

ID O9A956 PRELIMINARY; PRT; 613 AA.

AC O9A956;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE TOMB-DEPENDENT RECEPTOR, PUTATIVE.
CN Cc1138.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21173698; Pubmed=11259647;
RA Nierman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA Deboy R.T., Dodson R.J., Durkin A.S., Gwin M.L., Haft D.H.,
RA Kolony J.F., Smit J., Craven M.B., Knout H., Shetty J., Berry K.,
RA Salberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL: AE005792; AAK23122.1; -.
DR TIGR: Cc1138; -.
RT

KW Receptor; Complete proteome.
SQ SEQUENCE 613 AA; 64411 MW; ED65208A3D82B97D CRC64;

Query Match 82.9%; Score 29; DB 2; Length 613;
Best Local Similarity 50.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxafxg 8
|||
Db 370 WRAAAYAG 377

RESULT 11

ID O9RRD0 PRELIMINARY; PRT; 206 AA.

AC O9RRD0;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 3-DEMEHYLUBIQUINONE-9-3-METHYLTRANSFERASE, PUTATIVE.
CN D92562.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI;
RX MEDLINE=20036896; Pubmed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwin M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Uterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans RI.";
RL Science 286:1571-1577(1999).
DR EMBL: AE002085; AAF12100.1; -.
DR TIGR: DR2562; -.
DR InterPro: IPR001601; Meth-transf.
DR InterPro: IPR000051; SAM_bind.
DR Transferase; Methyltransferase; Ubiquinone; Complete proteome.
SQ SEQUENCE 206 AA; 22186 MW; 6F63EL369E12D870 CRC64;

Query Match 80.0%; Score 28; DB 2; Length 206;
Best Local Similarity 50.0%; Pred. No. 1e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wxxxafxg 8
|||
Db 111 WRTASFDG 118

RESULT 12

ID O9QON6 PRELIMINARY; PRT; 479 AA.

AC O9QON6;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PUTATIVE APHID TRANSMISSION FACTOR (FRAGMENT).
OS sugarcane yellow leaf virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
OC Unassigned Luteoviridae.
OX NCBI_TaxID=94290;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A.
RA Moonan F., Molina J.J., Mirkov T.E.;
RT "Sugarcane yellow leaf virus is a new virus with a genome that has

RT Poliovirus, Luteovirus, and Enamovirus properties."
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF157029; ADD45687.1; -
 DR InterPro: IPR000893; Luteo_ORF6.
 DR InterPro: IPR002929; PLRV_ORF5.
 DR InterPro: IPR002965; P_r1ch_extensn.
 DR Pfam: PF01690; PLRV_ORF5; 1.
 DR PRINTS: PR00910; LVIRUSORF6.
 DR PRINTS: PR01217; PRICHEXTENSN.
 FT NON_TER
 SQ SEQUENCE 479 AA: 51676 MW: 6C1DBAA2BA5EEC9 CRC64:

Query Match
 Best Local Similarity 80.0%; Score 28; DB 12; Length 479;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
 DB 129 WRAYAVSG 136

RESULT 13
 ID 09JH75 PRELIMINARY; PRT: 479 AA.
 AC 09JH75:
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DE PUTATIVE APHID TRANSMISSION FACTOR (FRAGMENT).
 OS sugarcane yellow leaf virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
 OC Unassigned Luteoviridae.
 OX NCBI_TaxID=94290;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. CP65-357;
 RA Smith G.R.;
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. CP65-357;
 RA MEDLINE:20318675; PubMed:10859394;
 RT Smith G.R., Borg Z., Lockhart B.E.L., Breithwalte K.S., Gibbs M.J.;
 RT "Sugarcane yellow leaf virus: a novel member of the Luteoviridae that
 RT probably arose by inter-species recombination."
 RL J. Gen. Virol. 81:1865-1869(2000).
 DR EMBL: AJ249447; CAB75437.1; -
 DR InterPro: IPR000893; Luteo_ORF6.
 DR InterPro: IPR002929; PLRV_ORF5.
 DR InterPro: IPR002965; P_r1ch_extensn.
 DR Pfam: PF01690; PLRV_ORF5; 1.
 DR PRINTS: PR00910; LVIRUSORF6.
 DR PRINTS: PR01217; PRICHEXTENSN.
 FT NON_TER
 SQ SEQUENCE 479 AA: 51616 MW: AD3B98A1B658652F CRC64:

Query Match
 Best Local Similarity 80.0%; Score 28; DB 12; Length 479;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
 DB 129 WRAYAVSG 136

RESULT 14
 ID 09BRR6 PRELIMINARY; PRT: 497 AA.
 AC 09BRR6:
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE SIMILAR TO RIKEN CDNA 2610017G09 GENE.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=KIDNEY ADENOCARCINOMA;
 RA Strausberg R.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC006112; AAH06112.1; -
 SQ SEQUENCE 497 AA: 54088 MW: B758E977CDA88F8F CRC64:

Query Match
 Best Local Similarity 80.0%; Score 28; DB 4; Length 497;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
 DB 4 WRSAVAG 11

RESULT 15
 ID 004892 PRELIMINARY; PRT: 530 AA.
 AC 004892:
 DT 01-JUL-1997 (TREMBlrel. 04, Created)
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
 DE CYTOCHROME P450 LIKE_TBP (EC 1.14.14.1).
 GN CYPB.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; eusterids I; Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxID=4097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRIGHT YELLOW 2;
 RX MEDLINE:96404975; PubMed:8809115;
 RA Suglura M., Sakaki T., Yabusaki Y., Ohkawa H.;
 RT "Cloning and expression in Escherichia coli and Saccharomyces
 RT cerevisiae of a novel tobacco cytochrome P-450-like cDNA."
 RL Biochim. Biophys. Acta 1308:231-240(1996).
 DR EMBL: D64052; BAA10929.1; -
 DR Mendel: 23988; Nicta:3078;23988.
 KW Oxidoreductase.
 SQ SEQUENCE 530 AA: 58916 MW: 118E04A903A749CA CRC64:

Query Match
 Best Local Similarity 80.0%; Score 28; DB 10; Length 530;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
 DB 254 WRDFAVG 261

RESULT 16
 ID 09BU02 PRELIMINARY; PRT: 59 AA.
 AC 09BU02:
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE SELENOPHOSPHATE SYNTHETASE 2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;

RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=RHABDOMYOSARCOMA;
 RA Strausberg R.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC002381; AA02381.1; -
 SQ SEQUENCE 59 AA; 5897 MW; 8DFC9070558436D5 CRC64;

Query Match
 Best Local Similarity 77.1%; Score 27; DB 4; Length 59;
 Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
 Db 49 WRLTGFSG 56

RESULT 17
 Q30824 PRELIMINARY; PRT; 64 AA.

ID Q30824
 AC Q30824
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE MHC CLASS I ANTIGEN (FRAGMENT).
 OS Ovis aries (Sheep).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae;
 OC Bovidae; Caprinae; Ovis.
 NC NCBITaxid=9940;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LATXA BREED: TISSUE=Blood;
 RA Vicario A., Jugo B., Martinez N., Santiago A., Aguirre A.I.,
 RA Mazon L.I., Estomba A.,
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U26008; AA67310.1; -
 DR InterPro: IPR001039; MHC_I.
 DR Pfam: PF00129; MHC_I; 1.
 DR ProDom: PD000050; MHC_I; 1.
 KW MHC.
 FT NON_TER 1
 FT NON_TER 64
 SQ SEQUENCE 64 AA; 7333 MW; 5702876FA3D10259 CRC64;

Query Match
 Best Local Similarity 77.1%; Score 27; DB 7; Length 64;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
 Db 56 WRASAMSG 63

RESULT 18
 Q9EYF4 PRELIMINARY; PRT; 247 AA.

ID Q9EYF4
 AC Q9EYF4
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE TAIL ASSEMBLY PROTEIN K.
 GN K.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia
 NC NCBITaxid=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7;
 RX PubMed=11111050;

RA Yokoyama K., Makino K., Kubota Y., Watanabe M., Kimura S.,
 RA Yutsudo C.H., Kurokawa K., Ishii K., Hattori M., Tatsuno I., Abe H.,
 RA Yon M., Iida T., Ohnishi M., Hayashi T., Yasunaga T., Honda T.,
 RA Sasakawa C., Shinagawa H.;
 RT "Complete nucleotide sequence of the prophage ϕ 1-Sakai carrying the
 RT Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli O157:H7
 RT strain derived from the Sakai outbreak.";
 RL Gene 258:127-139(2000)

DR EMBL: AP000400; BAB19565.1; -
 DR InterPro: IPR000555; MOV34.
 DR InterPro: IPR000064; NUPC_P60.
 DR Pfam: PF00877; NUPC_P60; 1.
 DR SMART: SM00232; JAB_MPN; 1.
 SQ SEQUENCE 247 AA; 28045 MW; 118570F99B48DA48 CRC64;

Query Match
 Best Local Similarity 77.1%; Score 27; DB 2; Length 247;
 Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
 Db 228 WHASAFSG 235

RESULT 19
 O59399 PRELIMINARY; PRT; 369 AA.

ID O59399
 AC O59399
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE FORMALDEHYDE DEHYDROGENASE (EC 1.2.1.1) (FORMALDEHYDE DEHYDROGENASE
 DE (GLUTATHIONE)) (FORMIC DEHYDROGENASE).
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 NC NCBITaxid=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CLINICAL ISOLATE;
 RX MEDLINE=97046207; PubMed=8891129;
 RA Kuemmerle N., Feucht H., Kaulfers P.M.;
 RT "Plasmid-mediated formaldehyde resistance in Escherichia coli:
 RT characterization of resistance gene.";
 RL Antimicrob. Agents Chemother. 40:2276-2279(1996).
 CC -1- CATALYTIC ACTIVITY: FORMALDEHYDE + NAD(+) = S-
 CC -1- FORMALYLGLUTATHIONE + NADH.
 CC -1- COFACTOR: REDUCED ZINC FOR ITS ACTIVITY (BY SIMILARITY).
 CC -1- SIMILARITY: TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE FAMILY.
 DR EMBL: X73835; CAA52057.1; -
 DR HSSP: P11766; 1TEH.
 DR InterPro: IPR002085; Adh_zinc.
 DR InterPro: IPR002328; Adh_zinc.
 DR Pfam: PF00107; adh_zinc; 1.
 DR PROSITE: PS00059; ADH_ZINC; 1.
 KW Oxidoreductase; Zinc.
 SQ SEQUENCE 369 AA; 39089 MW; C7A1FF5EBF636C6D CRC64;

Query Match
 Best Local Similarity 77.1%; Score 27; DB 2; Length 369;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
 Db 309 WKSARFG 316

RESULT 20
 O24687 PRELIMINARY; PRT; 369 AA.
 ID O24687
 AC O24687;

DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
 DE GLUTATHIONE DEPENDENT FORMALDEHYDE DEHYDROGENASE (EC 1.2.1.1).
 GN GDFALDH.
 OS Anabaena azollae.
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
 OX NCBI_TaxID-1164;
 RN 11
 RP SEQUENCE FROM N.A.
 RA Shaw W.H., Muthhead A., Arjol T., Plazinski J.;
 RL Submitted (Feb-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).
 CC -1- SIMILARITY: TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE FAMILY.
 DR EMBL: U89767; AAB61921.1; -.
 DR HSSP: P11766; 1TEH.
 DR InterPro: IPR002085; Adh_zinc.
 DR InterPro: IPR002328; Adh_zinc.
 DR InterPro: IPR000205; NAD_binding.
 DR Pfam: PF00107; adh_zinc; 1.
 DR PROSITE: PS00059; Adh_zinc; 1.
 KW Oxidoreductase; Zinc.
 SO SEQUENCE 369 AA; 39424 MW; 5E9F243D1024F594 CRC64;

Query Match 77.1%; Score 27; DB 2; Length 369;
 Best Local Similarity 50.0%; Pred. No. 3e+02;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
 I: | | |
 Db 309 WKGSATFG 316

Search completed: January 14, 2002, 07:39:49
 Job time: 960 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:36:01 ; Search time 103.51 Seconds
(without alignments)
5.725 Million cell updates/sec

Title: 09-185908-1h
Perfect score: 35
Sequence: 1 wxrxfxg 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database :

A.Ceneseq_1101: *
1: /SIDs2/gcgdata/geneseq/geneseqp/AA1980.DAT: *
2: /SIDs2/gcgdata/geneseq/geneseqp/AA1981.DAT: *
3: /SIDs2/gcgdata/geneseq/geneseqp/AA1982.DAT: *
4: /SIDs2/gcgdata/geneseq/geneseqp/AA1983.DAT: *
5: /SIDs2/gcgdata/geneseq/geneseqp/AA1984.DAT: *
6: /SIDs2/gcgdata/geneseq/geneseqp/AA1985.DAT: *
7: /SIDs2/gcgdata/geneseq/geneseqp/AA1986.DAT: *
8: /SIDs2/gcgdata/geneseq/geneseqp/AA1987.DAT: *
9: /SIDs2/gcgdata/geneseq/geneseqp/AA1988.DAT: *
10: /SIDs2/gcgdata/geneseq/geneseqp/AA1989.DAT: *
11: /SIDs2/gcgdata/geneseq/geneseqp/AA1990.DAT: *
12: /SIDs2/gcgdata/geneseq/geneseqp/AA1991.DAT: *
13: /SIDs2/gcgdata/geneseq/geneseqp/AA1992.DAT: *
14: /SIDs2/gcgdata/geneseq/geneseqp/AA1993.DAT: *
15: /SIDs2/gcgdata/geneseq/geneseqp/AA1994.DAT: *
16: /SIDs2/gcgdata/geneseq/geneseqp/AA1995.DAT: *
17: /SIDs2/gcgdata/geneseq/geneseqp/AA1996.DAT: *
18: /SIDs2/gcgdata/geneseq/geneseqp/AA1997.DAT: *
19: /SIDs2/gcgdata/geneseq/geneseqp/AA1998.DAT: *
20: /SIDs2/gcgdata/geneseq/geneseqp/AA1999.DAT: *
21: /SIDs2/gcgdata/geneseq/geneseqp/AA2000.DAT: *
22: /SIDs2/gcgdata/geneseq/geneseqp/AA2001.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	30	85.7	8	21	AAB06583	Claudin-3 cell adh
2	30	85.7	8	21	AAB06536	Claudin-3 cyclic c
3	30	85.7	8	21	AAB06645	Claudin-4 cell adh
4	30	85.7	8	21	AAB06598	Claudin-4 cyclic c
5	30	85.7	10	21	AAB06592	Claudin-3 cyclic c
6	30	85.7	10	21	AAB06601	Claudin-3 cyclic c
7	30	85.7	10	21	AAB06610	Claudin-3 cyclic c
8	30	85.7	10	21	AAB06619	Claudin-3 cyclic c
9	30	85.7	10	21	AAB06628	Claudin-3 cyclic c
10	30	85.7	10	21	AAB06654	Claudin-4 cyclic c
11	30	85.7	10	21	AAB06663	Claudin-4 cyclic c

12	30	85.7	10	21	AAB06672	Claudin-4 cyclic c
13	30	85.7	10	21	AAB06681	Claudin-4 cyclic c
14	30	85.7	10	21	AAB06690	Claudin-4 cyclic c
15	30	85.7	202	20	AAW88629	Secreted protein e
16	30	85.7	208	21	AAW52100	Gene 48 human secr
17	30	85.7	209	21	AAW43133	Human OREF ORF2897
18	30	85.7	210	20	AAW63307	Kidney injury asso
19	30	85.7	210	22	AAW75467	Human colon cancer
20	30	85.7	219	21	AAW51679	Murine clodin 3 pr
21	30	85.7	220	20	AAW13939	Human transmembran
22	28	80.0	59	18	AAW20137	H. pylori cytoplasm
23	28	80.0	379	21	AAW44794	P. pastoris Formal
24	28	80.0	392	21	AAW44795	Amino acid sequenc
25	28	80.0	417	21	AAW07679	Human AFP protein
26	28	80.0	496	22	AAW88466	Human membrane or
27	28	80.0	496	22	AAW88466	Hydrophobic domain
28	28	80.0	497	21	AAW12138	Claudin-6/9 cell a
29	27	77.1	8	21	AAB06764	Claudin-6/9 cyclic
30	27	77.1	8	21	AAB06814	Claudin cell adhes
31	27	77.1	8	21	AAB06916	Claudin-6/9 cyclic
32	27	77.1	10	21	AAB06772	Claudin-6/9 cyclic
33	27	77.1	10	21	AAB06781	Claudin-6/9 cyclic
34	27	77.1	10	21	AAB06789	Claudin-6/9 cyclic
35	27	77.1	10	21	AAB06798	Claudin-6/9 cyclic
36	27	77.1	10	21	AAB06806	Claudin-6/9 cyclic
37	27	77.1	78	21	AAB43420	Human cancer assoc
38	27	77.1	126	21	AAB24453	Human secreted pro
39	27	77.1	208	22	AAB95797	Human protein sequ
40	27	77.1	215	21	AAB24485	Human secreted pro
41	27	77.1	217	22	AAB64401	Murine clodin 6 pr
42	27	77.1	219	21	AAW51681	Human Prol1488 (UNQ
43	27	77.1	220	21	AAW99434	Human polypeptide
44	27	77.1	220	22	AAW39937	Human membrane or
45	27	77.1	220	22	AAW88319	Protein of the inv
46	27	77.1	220	22	AAB65183	Human Prol1488 prot
47	27	77.1	220	22	AAW50968	Human polypeptide
48	27	77.1	235	22	AAW41723	Arabidopsis thaliana
49	27	77.1	254	21	AAW17084	Arabidopsis thaliana
50	27	77.1	255	21	AAW16747	Arabidopsis thaliana

ALIGNMENTS

RESULT 1
AAB06583 standard: peptide; 8 AA.
XX
AC AAB06583;
XX
DT 28-SEP-2000 (first entry)
XX
DE Claudin-3 cell adhesion recognition sequence SEQ ID NO: 58.
XX
KW Claudin-3 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
XX graft rejection.
XX
XX Mammalia.
OS
XX
PN WO200026360-A1.
XX
PD 11-MAY-2000.
XX
PE 03-NOV-1999; 99WO-CA01029.
XX
PR 03-NOV-1998; 98US-0185908.
XX
PR 30-MAR-1999; 99US-0282029.
XX
PA (ADHE-) ADHEREX TECHNOLOGIES INC.
XX
PI Blaschuck OW, Symonds JM, Gour BJ;
XX

DR WPI: 2000-365610/31.
XX Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
XX
PS Claim 52; Page 99; 121pp; English.
CC The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-3 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site.
XX
SQ Sequence 8 AA;

Query Match 85.7%; Score 30; DB 21; Length 8;
Best Local Similarity 62.5%; Pred. No. 4.3e+05;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
|| || |
Db 1 wrvsafifg 8

RESULT 2
AAB06636
ID AAB06636 standard; peptide; 8 AA.
XX
AC AAB06636;
XX
DT 28-SEP-2000 (first entry)
XX
DE Claudin-3 cyclic cell adhesion recognition sequence SEQ ID NO: 269.
XX
XX Claudin-3 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection; cyclic.
XX
OS Mammalia.
XX
XX WO200026360-A1.
PN
XX
PD 11-MAY-2000.
XX
PE 03-NOV-1999; 99WO-CA01029.
XX
PR 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
XX
PA (ADHE-) ADHEREX TECHNOLOGIES INC.
XX
PI Blaschuck OW, Symonds JM, Gour BJ;
XX WPI: 2000-365610/31.
DR
XX
PT Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
XX
PS Claim 55; Page 100; 121pp; English.
XX
XX The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-3 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery

CC to the desired target site. The present sequence has a cyclic
CC conformation.
XX
SQ Sequence 8 AA;

Query Match 85.7%; Score 30; DB 21; Length 8;
Best Local Similarity 62.5%; Pred. No. 4.3e+05;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
|| || |
Db 1 wrvsafifg 8

RESULT 3
AAB06645
ID AAB06645 standard; peptide; 8 AA.
XX
AC AAB06645;
XX
DT 28-SEP-2000 (first entry)
XX
DE Claudin-4 cell adhesion recognition sequence SEQ ID NO: 50.
XX
XX Claudin-4 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection.
XX
OS Mammalia.
XX
XX WO200026360-A1.
PN
XX
PD 11-MAY-2000.
XX
PE 03-NOV-1999; 99WO-CA01029.
XX
PR 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
XX
PA (ADHE-) ADHEREX TECHNOLOGIES INC.
XX
PI Blaschuck OW, Symonds JM, Gour BJ;
XX WPI: 2000-365610/31.
DR
XX
PT Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
XX
PS Claim 58; Page 100; 121pp; English.
XX
XX The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-4 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site.
XX
SQ Sequence 8 AA;

Query Match 85.7%; Score 30; DB 21; Length 8;
Best Local Similarity 62.5%; Pred. No. 4.3e+05;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
|| || |
Db 1 wrvsafifg 8

RESULT 4
AAB06598 ID AAB06598 standard; peptide: 8 AA.
XX AC AAB06598;
XX DT 28-SEP-2000 (first entry)
XX DE Claudin-4 cyclic cell adhesion recognition sequence SEQ ID NO: 216.
XX KM Claudin-4 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KM graft rejection; cyclic.
XX OS Mammalia.
XX PN WO200026360-A1.
XX PD 11-MAY-2000.
XX PF 03-NOV-1999; 99WO-CA01029.
XX PR 03-NOV-1998; 98US-0185908.
XX PR 30-MAR-1999; 99US-0282029.
XX PA (ADHE-) ADHEREX TECHNOLOGIES INC.
XX PI Blaschuck OW, Symonds JM, Gour BJ;
XX DR WPI: 2000-365610/31.
XX PT Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX PS Claim 61; Page 101; 121pp; English.
XX CC The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-4 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.
XX SQ Sequence 8 AA:

Query Match 85.7%; Score 30; DB 21; Length 8;
Best Local Similarity 62.5%; Pred. No. 4.3e+05;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
 || || |
DB 1 wrvlatfg 8

OS Mammalia.
XX PN WO200026360-A1.
XX PD 11-MAY-2000.
XX PF 03-NOV-1999; 99WO-CA01029.
XX PR 03-NOV-1998; 98US-0185908.
XX PR 30-MAR-1999; 99US-0282029.
XX PA (ADHE-) ADHEREX TECHNOLOGIES INC.
XX PI Blaschuck OW, Symonds JM, Gour BJ;
XX DR WPI: 2000-365610/31.
XX PT Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX PS Claim 55; Page 99; 121pp; English.
XX CC The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-3 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.
XX SQ Sequence 10 AA:

Query Match 85.7%; Score 30; DB 21; Length 10;
Best Local Similarity 62.5%; Pred. No. 3;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
 || || |
DB 2 wrvsafig 9

RESULT 6
AAB06601 ID AAB06601 standard; peptide: 10 AA.
XX AC AAB06601;
XX DT 28-SEP-2000 (first entry)
XX DE Claudin-3 cyclic cell adhesion recognition sequence SEQ ID NO: 234.
XX KM Claudin-3 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection; cyclic.
XX OS Mammalia.
XX PN WO200026360-A1.
XX PD 11-MAY-2000.
XX PF 03-NOV-1999; 99WO-CA01029.
XX PR 03-NOV-1998; 98US-0185908.
XX PR 30-MAR-1999; 99US-0282029.
XX PA (ADHE-) ADHEREX TECHNOLOGIES INC.
XX PI Blaschuck OW, Symonds JM, Gour BJ;

XX WPI: 2000-365610/31.
DR
XX Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
PS Claim 55; Page 99; 121pp; English.
XX
CC The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-3 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.
XX
SQ Sequence 10 AA;

Query Match 85.7%; Score 30; DB 21; Length 10;
Best Local Similarity 62.5%; Pred. No. 3;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 wrxxafxg 8
| | | | |
Db 2 wrvsatfg 9

RESULT 7
AAB06610
ID AAB06610 standard; peptide; 10 AA.
XX
AC AAB06610;
XX
DT 28-SEP-2000 (first entry)
XX
DE Claudin-3 cyclic cell adhesion recognition sequence SEQ ID NO: 243.
XX
KW Claudin-3 modulating agent; cell adhesion recognition sequence;
KM CAR sequence; autoimmune disease; inflammatory disease; cancer;
KM graft rejection; cyclic.
XX
OS Mammalia.
XX
PN WO200026360-A1.
XX
PD 11-MAY-2000.
XX
PF 03-NOV-1999; 99WO-CA01029.
XX
PR 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
XX
PA (ADHE-) ADHEREX TECHNOLOGIES INC.
XX
PI Blaschuck OW, Symonds JM, Gour BJ;
XX
DR WPI: 2000-365610/31.
XX
PT Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
PS Claim 55; Page 99; 121pp; English.
XX
CC The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-3 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,

CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.
XX
SQ Sequence 10 AA;

Query Match 85.7%; Score 30; DB 21; Length 10;
Best Local Similarity 62.5%; Pred. No. 3;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 wrxxafxg 8
| | | | |
Db 2 wrvsatfg 9

RESULT 8
AAB06619
ID AAB06619 standard; peptide; 10 AA.
XX
AC AAB06619;
XX
DT 28-SEP-2000 (first entry)
XX
DE Claudin-3 cyclic cell adhesion recognition sequence SEQ ID NO: 252.
XX
KW Claudin-3 modulating agent; cell adhesion recognition sequence;
KM CAR sequence; autoimmune disease; inflammatory disease; cancer;
KM graft rejection; cyclic.
XX
OS Mammalia.
XX
PN WO200026360-A1.
XX
PD 11-MAY-2000.
XX
PF 03-NOV-1999; 99WO-CA01029.
XX
PR 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
XX
PA (ADHE-) ADHEREX TECHNOLOGIES INC.
XX
PI Blaschuck OW, Symonds JM, Gour BJ;
XX
DR WPI: 2000-365610/31.
XX
PT Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
PS Claim 55; Page 100; 121pp; English.
XX
CC The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-3 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.
XX
SQ Sequence 10 AA;

Query Match 85.7%; Score 30; DB 21; Length 10;
Best Local Similarity 62.5%; Pred. No. 3;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 wrxxafxg 8
| | | | |

DB	2	wrvsaflg 9	
RESULT 9			
AB06628			
ID	AB06628	standard; peptide; 10 AA.	
XX			
AC	AB06628;		
XX			
DT	28-SEP-2000	(first entry)	
XX			
DE	Claudin-3	cyclic cell adhesion recognition sequence SEQ ID NO: 261.	
XX			
KW	Claudin-3	modulating agent; cell adhesion recognition sequence;	
KW	CAR sequence;	autoimmune disease; inflammatory disease; cancer;	
KW	graft rejection;	cyclic.	
XX			
OS	Mammalia.		
XX			
PN	WO200026360-A1.		
XX			
PD	11-MAY-2000.		
XX			
PF	03-NOV-1999;	99WO-CA01029.	
XX			
PR	03-NOV-1998;	98US-0185908.	
PR	30-MAR-1999;	99US-0282029.	
PA	(ADHE-) ADHEREX TECHNOLOGIES INC.		
XX			
PI	Blaschuck OW, Symonds JM, Gour BJ:		
XX			
DR	WPI: 2000-365610/31.		
XX			
PT	Antibody modulation of claudin-mediated cell adhesion for increasing		
PT	vasopermeability, for delivering drugs to tumors and the nervous system		
PT	and across the skin -		
XX			
PS	Claim 55; Page 100; 121pp; English.		
XX			
CC	The present invention relates to the use of peptides as claudin-mediated		
CC	cell adhesion modulators. The claudin-3 group of proteins are cadherins,		
CC	which are membrane glycoproteins involved in cell adhesion. In some		
CC	situations, cell adhesion occurs at abnormal levels, and these peptides		
CC	can be used to modulate these levels, and thus treat autoimmune diseases,		
CC	inflammatory diseases and cancer, and aid wound healing and implant		
CC	adhesion. In addition, they can also be used to facilitate drug delivery		
CC	to the desired target site. The present sequence has a cyclic		
CC	conformation.		
XX			
SQ	Sequence 10 AA;		
Query Match	85.7%;	Score 30;	DB 21; Length 10;
Best Local Similarity	62.5%;	Pred. No. 3;	
Matches 5; Conservative	0;	Mismatches 3;	Indels 0; Gaps 0;
OY	1	wrxaxfxg 8	
DB	2	wrvsaflg 9	
RESULT 10			
AB06654			
ID	AB06654	standard; peptide; 10 AA.	
XX			
AC	AB06654;		
XX			
DT	28-SEP-2000	(first entry)	
XX			
DE	Claudin-4	cyclic cell adhesion recognition sequence SEQ ID NO: 172.	
XX			
KW	Claudin-4	modulating agent; cell adhesion recognition sequence;	

KW	CAR sequence;	autoimmune disease; inflammatory disease; cancer;	
KW	graft rejection;	cyclic.	
XX			
OS	Mammalia.		
XX			
PN	WO200026360-A1.		
XX			
PD	11-MAY-2000.		
XX			
PF	03-NOV-1999;	99WO-CA01029.	
XX			
PR	03-NOV-1998;	98US-0185908.	
PR	30-MAR-1999;	99US-0282029.	
PA	(ADHE-) ADHEREX TECHNOLOGIES INC.		
XX			
PI	Blaschuck OW, Symonds JM, Gour BJ:		
XX			
DR	WPI: 2000-365610/31.		
XX			
PT	Antibody modulation of claudin-mediated cell adhesion for increasing		
PT	vasopermeability, for delivering drugs to tumors and the nervous system		
PT	and across the skin -		
XX			
PS	Claim 61; Page 101; 121pp; English.		
XX			
CC	The present invention relates to the use of peptides as claudin-mediated		
CC	cell adhesion modulators. The claudin-4 group of proteins are cadherins,		
CC	which are membrane glycoproteins involved in cell adhesion. In some		
CC	situations, cell adhesion occurs at abnormal levels, and these peptides		
CC	can be used to modulate these levels, and thus treat autoimmune diseases,		
CC	inflammatory diseases and cancer, and aid wound healing and implant		
CC	adhesion. In addition, they can also be used to facilitate drug delivery		
CC	to the desired target site. The present sequence has a cyclic		
CC	conformation.		
XX			
SQ	Sequence 10 AA;		
Query Match	85.7%;	Score 30;	DB 21; Length 10;
Best Local Similarity	62.5%;	Pred. No. 3;	
Matches 5; Conservative	0;	Mismatches 3;	Indels 0; Gaps 0;
OY	1	wrxaxfxg 8	
DB	2	wrvsaflg 9	
RESULT 11			
AB06663			
ID	AB06663	standard; peptide; 10 AA.	
XX			
AC	AB06663;		
XX			
DT	28-SEP-2000	(first entry)	
XX			
DE	Claudin-4	cyclic cell adhesion recognition sequence SEQ ID NO: 181.	
XX			
KW	Claudin-4	modulating agent; cell adhesion recognition sequence;	
KW	CAR sequence;	autoimmune disease; inflammatory disease; cancer;	
KW	graft rejection;	cyclic.	
XX			
OS	Mammalia.		
XX			
PN	WO200026360-A1.		
XX			
PD	11-MAY-2000.		
XX			
PF	03-NOV-1999;	99WO-CA01029.	
XX			
PR	03-NOV-1998;	98US-0185908.	
PR	30-MAR-1999;	99US-0282029.	
XX			

PA (ADHE-) ADHEREX TECHNOLOGIES INC.
XX Blaschuck OW, Symonds JM, Gour BJ;
XX WPI; 2000-365610/31.
DR
XX Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
XX
PS Claim 61; Page 101; 121pp; English.
XX
XX The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-4 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.
XX
SQ Sequence 10 AA:

Query Match 85.7%; Score 30; DB 21; Length 10;
Best Local Similarity 62.5%; Pred. No. 3;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxxafxg 8
|| || |
Db 2 wrvtafig 9

RESULT 12
AAB06672
ID AAB06672 standard; peptide; 10 AA.
XX
XX AAB06672;
AC
XX 28-SEP-2000 (first entry)
DT
XX
XX Claudin-4 cyclic cell adhesion recognition sequence SEQ ID NO: 190.
DE
XX
XX Claudin-4 modulating agent; cell adhesion recognition sequence;
KM CAR sequence; autoimmune disease; inflammatory disease; cancer;
KM graft rejection; cyclic.
XX
XX Mammalia.
OS
XX WO200026360-A1.
PN
XX 11-MAY-2000.
PD
XX
XX 03-NOV-1999; 99WO-CA01029.
PE
XX 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
XX
XX (ADHE-) ADHEREX TECHNOLOGIES INC.
PA
XX Blaschuck OW, Symonds JM, Gour BJ;
XX WPI; 2000-365610/31.
DR
XX
XX Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
XX Claim 61; Page 101; 121pp; English.
XX
XX The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-4 group of proteins are cadherins,

CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.
XX
SQ Sequence 10 AA:

Query Match 85.7%; Score 30; DB 21; Length 10;
Best Local Similarity 62.5%; Pred. No. 3;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 wxxxafxg 8
|| || |
Db 2 wrvtafig 9

RESULT 13
AAB06681
ID AAB06681 standard; peptide; 10 AA.
XX
XX AAB06681;
AC
XX 28-SEP-2000 (first entry)
DT
XX
XX Claudin-4 cyclic cell adhesion recognition sequence SEQ ID NO: 199.
DE
XX
XX Claudin-4 modulating agent; cell adhesion recognition sequence;
KM CAR sequence; autoimmune disease; inflammatory disease; cancer;
KM graft rejection; cyclic.
XX
XX Mammalia.
OS
XX WO200026360-A1.
PN
XX 11-MAY-2000.
PD
XX
XX 03-NOV-1999; 99WO-CA01029.
PE
XX 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
XX
XX (ADHE-) ADHEREX TECHNOLOGIES INC.
PA
XX Blaschuck OW, Symonds JM, Gour BJ;
XX WPI; 2000-365610/31.
DR
XX
XX Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
XX Claim 61; Page 101; 121pp; English.
XX
XX The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-4 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.
XX
SQ Sequence 10 AA:

Query Match 85.7%; Score 30; DB 21; Length 10;
Best Local Similarity 62.5%; Pred. No. 3;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1
OY 1 wrxxafxg 8
11 111
DB 2 wrvtafig 9

RESULT 14
AAB06690 standard; peptide: 10 AA.
XX
AC AAB06690;
XX
DT 28-SEP-2000 (first entry)
XX
DE Claudin-4 cyclic cell adhesion recognition sequence SEQ ID NO: 208.
XX
KW Claudin-4 modulating agent; cell adhesion recognition sequence;
KW CAR sequence; autoimmune disease; inflammatory disease; cancer;
KW graft rejection; cyclic.
XX
OS Mammalia.
XX
PN WO200026360-A1.
XX
PD 11-MAY-2000.
XX
PF 03-NOV-1999; 99WO-CA01029.
XX
PR 03-NOV-1998; 98US-0185908.
PR 30-MAR-1999; 99US-0282029.
XX
PA (ADHE-) ADHEREX TECHNOLOGIES INC.
XX
PI Blaschuck OM, Symonds JM, Gour BJ;
XX
DR WPI; 2000-365610/31.
XX
PT Antibody modulation of claudin-mediated cell adhesion for increasing
PT vasopermeability, for delivering drugs to tumors and the nervous system
PT and across the skin -
XX
PS Claim 61; Page 101; 121pp; English.
XX
CC The present invention relates to the use of peptides as claudin-mediated
CC cell adhesion modulators. The claudin-4 group of proteins are cadherins,
CC which are membrane glycoproteins involved in cell adhesion. In some
CC situations, cell adhesion occurs at abnormal levels, and these peptides
CC can be used to modulate these levels, and thus treat autoimmune diseases,
CC inflammatory diseases and cancer, and aid wound healing and implant
CC adhesion. In addition, they can also be used to facilitate drug delivery
CC to the desired target site. The present sequence has a cyclic
CC conformation.
XX
SQ Sequence 10 AA:

DE Secreted protein encoded by gene 96 clone HRDFB85.
XX
XX Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
PN WO9854963-A2.
XX
PD 10-DEC-1998.
XX
PE 04-JUN-1998; 98WO-US11422.
XX
XX 18-DEC-1997; 97US-0070923.
PR 06-JUN-1997; 97US-0048877.
PR 06-JUN-1997; 97US-0048881.
PR 06-JUN-1997; 97US-0048884.
PR 06-JUN-1997; 97US-0048893.
PR 06-JUN-1997; 97US-0048896.
PR 06-JUN-1997; 97US-0048899.
PR 06-JUN-1997; 97US-0048915.
PR 06-JUN-1997; 97US-0048949.
PR 06-JUN-1997; 97US-0048964.
PR 06-JUN-1997; 97US-0048972.
PR 06-JUN-1997; 97US-0049020.
PR 06-JUN-1997; 97US-0049375.
PR 05-SEP-1997; 97US-0057628.
PR 05-SEP-1997; 97US-0057635.
PR 05-SEP-1997; 97US-0057644.
PR 05-SEP-1997; 97US-0057647.
PR 05-SEP-1997; 97US-0057650.
PR 05-SEP-1997; 97US-0057661.
PR 05-SEP-1997; 97US-0057667.
PR 05-SEP-1997; 97US-0057761.
PR 05-SEP-1997; 97US-0057764.
PR 05-SEP-1997; 97US-0057770.
PR 05-SEP-1997; 97US-0057775.
PR 05-SEP-1997; 97US-0057778.
PR 06-JUN-1997; 97US-0048875.
PR 06-JUN-1997; 97US-0048878.
PR 06-JUN-1997; 97US-0048882.
PR 06-JUN-1997; 97US-0048885.
PR 06-JUN-1997; 97US-0048894.
PR 06-JUN-1997; 97US-0048897.
PR 06-JUN-1997; 97US-0048900.
PR 06-JUN-1997; 97US-0048916.
PR 06-JUN-1997; 97US-0048962.
PR 06-JUN-1997; 97US-0048970.
PR 06-JUN-1997; 97US-0048974.
PR 05-SEP-1997; 97US-0049373.
PR 05-SEP-1997; 97US-0057584.
PR 05-SEP-1997; 97US-0057629.
PR 05-SEP-1997; 97US-0057642.
PR 05-SEP-1997; 97US-0057645.
PR 05-SEP-1997; 97US-0057648.
PR 05-SEP-1997; 97US-0057651.
PR 05-SEP-1997; 97US-0057662.
PR 05-SEP-1997; 97US-0057668.
PR 05-SEP-1997; 97US-0057762.
PR 05-SEP-1997; 97US-0057765.
PR 05-SEP-1997; 97US-0057771.
PR 05-SEP-1997; 97US-0057776.
PR 06-JUN-1997; 97US-0048876.
PR 06-JUN-1997; 97US-0048880.
PR 06-JUN-1997; 97US-0048883.
PR 06-JUN-1997; 97US-0048892.
PR 06-JUN-1997; 97US-0048895.

PR 06-JUN-1997; 97US-0048898.
PR 06-JUN-1997; 97US-0048901.
PR 06-JUN-1997; 97US-0048917.
PR 06-JUN-1997; 97US-0048963.
PR 06-JUN-1997; 97US-0048971.
PR 06-JUN-1997; 97US-0049019.
PR 06-JUN-1997; 97US-0049374.
PR 05-SEP-1997; 97US-0057627.
PR 05-SEP-1997; 97US-0057634.
PR 05-SEP-1997; 97US-0057646.
PR 05-SEP-1997; 97US-0057649.
PR 05-SEP-1997; 97US-0057654.
PR 05-SEP-1997; 97US-0057666.
PR 05-SEP-1997; 97US-0057760.
PR 05-SEP-1997; 97US-0057763.
PR 05-SEP-1997; 97US-0057769.
PR 05-SEP-1997; 97US-0057774.
PR 05-SEP-1997; 97US-0057777.
XX

XX (HUMA-) HUMAN GENOME SCI INC.

PI Brewer LA, Carter KC, Dillon PJ, Edner R, Endress GA;
PI Fan P, Feng P, Ferrie AM, Fischer CL, Florence C;
PI Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW;
PI Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM;
PI Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;
XX WPI: 1999-059865/05.
DR N-PSDB; AAV84506.
XX

PT New isolated human genes and the secreted polypeptides they encode -
PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders

XX Claim 11; Page 533; 772pp; English.

XX The invention relates to nucleic acid sequences (AAV84411 to AAV84633)
CC encoding human secreted proteins (AAW88534 to AAW88756). The secreted
CC protein gene sequences are deposited with the ATCC under deposit numbers
CC AACC 97978, 97974, 97975, 97976, 97977, 209007, 209008, 209009, 209010,
CC 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host
CC cells comprising recombinant vectors containing the nucleic acid
CC sequences are used for the recombinant production of the secreted
CC proteins. The polynucleotide and amino acid sequences are useful for are
CC useful for preventing, treating or ameliorating medical conditions e.g.
CC by protein or gene therapy. Pathological conditions can be also
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the polynucleotides, based on
CC which tissues they are most highly expressed in, and include developing
CC products for the diagnosis or treatment of cancer, neurodegenerative
CC disorders, developmental abnormalities and foetal deficiencies, blood
CC disorders, tumours, leukemias, diseases of the immune system, autoimmune
CC diseases, hepatic and renal disease, lymphomas, inflammation, allergies,
CC ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,
CC reterositis, prostate diseases, obesity, disorders involving osteoclasts
CC such as osteoporosis, arthritis or malignancies, diseases of testes,
CC lung or thymus, digestive/endocrine disorders, infections and AIDS. The
CC polypeptides are also useful for identifying their binding partners.
CC The present sequence represents human secreted protein (see descriptor
CC line for gene number and clone identification).

XX Sequence 202 AA;

Query Match 85.7%; Score 30; DB 20; Length 202;

Best Local Similarity 62.5%; Pred. No. 49;

Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
II II II
DB 22 wrvtafig 29

RESULT 16

AAB52100 -
ID AAB52100 standard; Protein: 208 AA.

XX AAB52100;

XX 21-FEB-2001 (first entry)

DE Gene 48 human secreted protein homologous amino acid sequence #149.

XX Human; secreted protein; cytosolic; immunosuppressive; neutrotic;
XX neutrotrophic; antiviral; antiallergic; hepatotropic; antidiabetic;
XX antiinflammatory; antitumor; anticonvulsant; antibacterial;
XX antifungal; antiparasitic; cardiant; cancer; immune disease; allergy;
XX cardiovascular disorder; wound healing; infection; neurological disease.
OS

XX Mus musculus.

XX WO200061596-A1.

XX 19-OCT-2000.

XX 06-APR-2000; 2000WO-US08983.

XX 09-APR-1999; 99US-0128703.

XX 20-JAN-2000; 2000US-0176068.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (ROSE/) ROSEN C A.

XX Ruben SM, Komatsoulis G;

XX WPI: 2000-611665/58.

XX Fifty nucleic acid molecules encoding human secreted proteins, useful
PT in the prevention, treatment and diagnosis of cancer, immune disorders,
PT cardiovascular disorders and neurological diseases -
XX
XX
XX Disclosure; Page 83; 505pp; English.

XX Polynucleotide sequences AAC95521 - AAC95570 represent cDNA encoding 50
CC human secreted proteins AAB52012 - AAB52161. Sequences AAB52062 -
CC AAB52103 represent alternative polypeptides encoded by the genes, and
CC amino acid sequences with which they share homology. The genes and
CC proteins have activities dependent on the tissues and cells in which they
CC are expressed. Examples of their activities include cytostatic;
CC immunosuppressive; neutrotic; neutrotrophic; antiviral; antiallergic;
CC hepatotropic; antidiabetic; antiinflammatory; antitumor; vulnary;
CC anticonvulsant; antibacterial; antifungal; antiparasitic; and cardiant.
CC The secreted proteins, polypeptides, antagonists and agonists may be
CC useful in treating, preventing and/or diagnosing diseases and disorders
CC such as cancer, particularly breast and ovarian cancer, and other cancers
CC of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract,
CC liver, lung, or urogenital. Immune disorders such as Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; cardiovascular disorders such as
CC myocardial ischaemia; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases such as viral,
CC bacterial, fungal and parasitic infections may also be treated using the
CC proteins and polynucleotides of the invention. Sequences AAC95512 -
CC AAC95520 and AAB52011 are used in the isolation and characterisation of
CC the proteins and polynucleotides of the invention.

XX Sequence 208 AA;

Query Match 85.7%; Score 30; DB 21; Length 208;

Best Local Similarity 62.5%; Pred. No. 51;

Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1wxrxfxg 8
11111
Db 30 wrvtafig 37

RESULT 17
AAB43133
ID AAB43133 standard; Protein: 209 AA.

XX AAB43133:
XX
DT 08-FEB-2001 (first entry)

XX Human ORFX ORF2897 polypeptide sequence SEQ ID NO:5794.
XX
XX Human: open reading frame; ORFX: detection; cytostatic; hepatotropic;
KW vulnery; antiparkinsonian; antiarthritic; immunosuppressant; cardiant;
KW anticonvulsant; osteopathic; antidiabetic;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatologic; immunosuppressive; antineoplastic;
KW antiviral; antibacterial; antifungal; antineumatic; antihydroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antineoplastic disease; coagulation;
KW thrombosis; contraceptive.

XX Homo sapiens.
XX
XX WO200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000MO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shimkets RA, Leach M;
PI
DR WPI: 2000-602362/57.
DR N-PSDB: AAC77342.

XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 11; Page 4961; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antiparkinsonian; antiarthritic; immunosuppressant;
CC osteopathic; anticonvulsant; antidiabetic; immunosuppressive;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatologic; immunosuppressive;
CC antineoplastic; antibacterial; antifungal; antineumatic;
CC antihydroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,

CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX Sequence 209 AA:

Query Match 85.7%; Score 30; DB 21; Length 209;
Best Local Similarity 62.5%; Pred. No. 51;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1wxrxfxg 8
11111
Db 30 wrvtafig 37

RESULT 18
AAB6307
ID AAB6307 standard; Protein: 210 AA.

XX AAB6307:
XX
AC AAB6307;
XX
XX 01-MAR-1999 (first entry)
XX
XX Kidney injury associated molecule HM014 protein.
DE
XX Kidney injury associated molecule; kidney injury related molecule;
KW kidney injury associated molecule; kidney injury related molecule;
KW kidney tissue growth promotion; regeneration; renal condition;
KW acute renal failure; acute nephritis; tumour.
XX
XX Rattus sp.
XX
XX WO9853071-A1.
XX
XX 26-NOV-1998.
XX
XX 22-MAY-1998; 98MO-US10547.
XX
XX 23-MAY-1997; 97US-0047491.
PR 23-MAY-1997; 97US-0047490.
XX
XX (BIOG) BIOGEN INC.
XX
XX Cate RL, Hession CA, Sanicola-Nadel M, Wei H;
PI
XX WPI: 1999-045312/04.
DR N-PSDB: AAV80582.

XX Kidney injury-associated molecule, KIM, polypeptides - upregulated
PT in injured or regenerating tissues, useful to promote tissue growth
PT and regeneration, especially to treat renal conditions
XX
XX Claim 17; Page 46-47; 213pp; English.

XX The present sequence represents a kidney injury associated molecule
CC (KIM) protein. KIM proteins can be administered therapeutically
CC by expressing KIM encoding polynucleotides, to promote growth and/or
CC survival of damaged tissue (e.g. renal tissue), since the KIM proteins
CC are upregulated in injured or regenerating (especially renal) tissues.
CC KIM fusion proteins, conjugates, antibodies and vectors can also be used
CC therapeutically, e.g. these or the KIM proteins may be included with an
CC acceptable carrier in pharmaceutical compositions, useful for therapy/
CC prophylaxis of conditions associated with dysfunction/dysregulation of
CC KIM genes or proteins, especially renal diseases or impairments of renal
CC function in humans (e.g. acute renal failure, acute nephritis). The
CC polynucleotides can be used to produce antisense sequences which, when
CC internalised into cells, can disrupt expression of a cellular KIM gene,
CC also useful in therapy (e.g. to block the growth of tumours dependent on
CC KIM for growth) or compositions. The proteins and polynucleotides are
CC useful diagnostically e.g. to detect and quantify renal injury/disease
CC (indicative of increased risk, or presence of, renal injury or impaired
CC function), or abnormal responses to tissue injury (indicative of

CC increased risk, or presence of, an autoimmune response or abnormal
CC tissue growth arising from/affecting renal tissue). The proteins can
CC also be used to locate KIM-producing cells (especially specific loci,
CC e.g. tissue masses abnormally producing/expressing KIM such as tumours
CC arising from/affecting renal tissue), by contacting cells with an
CC imageable KIM-binding reagent and imaging reagent accumulation.

XX Sequence 210 AA;

Query Match 85.7%; Score 30; DB 20; Length 210;

Best Local Similarity 62.5%; Pred. No. 51;

Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxafxg 8

DB 30 wrvtatfig 37

RESULT 19

AAAG75467 standard; Protein; 210 AA.

AC AAG75467;

DT 03-SEP-2001 (first entry)

DE Human colon cancer antigen protein SEQ ID NO:6231.

KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KM colorectal carcinoma.

OS Homo sapiens.

PN WO200122920-A2.

PD 05-APR-2001.

PF 28-SEP-2000; 2000WO-US26524.

PR 29-SEP-1999; 99US-0157137.

PR 03-NOV-1999; 99US-0163280.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Ruben SM, Barash SC, Birse CE, Rosen CA;

DR WPI; 2001-235357/24.

DR N-PSDB; AAH34872.

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX Claim 11; Page 7661; 9803pp; English.

AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patient's own production of P.
CC Additionally, N may be used to produce the colon cancer-associated PS,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAH77789 represent sequences used in the exemplification of the
CC present invention.

CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for

CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 210 AA;

Query Match 85.7%; Score 30; DB 22; Length 210;

Best Local Similarity 62.5%; Pred. No. 51;

Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxafxg 8

DB 31 wrvtatfig 38

RESULT 20

AAAY51679 standard; Protein; 219 AA.

AC AAY51679;

DT 02-JUN-2000 (first entry)

DE Murine clodrin 3 protein.

KW Clodrin 3; murine; tight junction-constituting membrane protein;

KM medicine.

OS Mus sp.

PN JP2000032984-A.

PD 02-FEB-2000.

PF 26-JUN-1998; 98JP-0179847.

PR 15-MAY-1998; 98JP-0133215.

PA (EISA) EISAI CO LTD.

DR WPI; 2000-285512/25.

DR N-PSDB; AAZ89151.

XX Tight junction-constituting membrane protein clodrin family - useful in
XX the medical field

PS Claim 3; Page 15; 22pp; Japanese.

XX This invention describes novel murine nucleic acid sequences encoding the
CC clodrin family of tight junction (TJ)-constituting membrane protein. The
CC membrane protein can be used in medical field. This sequence represents
CC the clodrin 3 protein described in the method of the invention.

XX Sequence 219 AA;

Query Match 85.7%; Score 30; DB 21; Length 219;

Best Local Similarity 62.5%; Pred. No. 53;

Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxafxg 8

DB 29 wrvtatfig 36

Search completed: January 14, 2002, 07:36:02
Job time: 813 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:23:45 ; Search time 48.05 Seconds
(without alignments)
3.747 Million cell updates/sec

Title: 09-185908-1h
Perfect score: 35
Sequence: 1 wrxxafxg 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues
Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCrus.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	27	77.1	376	1	US-08-608-241-2	Sequence 2, Appl
2	27	77.1	376	2	US-08-922-182-2	Sequence 2, Appl
3	27	77.1	376	2	US-08-919-953-2	Sequence 2, Appl
4	27	77.1	376	4	US-09-192-983-2	Sequence 2, Appl
5	26	74.3	462	4	US-09-352-159-6	Sequence 6, Appl
6	26	74.3	462	4	US-09-352-159-8	Sequence 8, Appl
7	26	74.3	462	4	US-09-352-168-6	Sequence 6, Appl
8	26	74.3	462	4	US-09-352-168-8	Sequence 8, Appl
9	26	74.3	463	4	US-09-352-159-11	Sequence 11, Appl
10	26	74.3	463	4	US-09-352-168-11	Sequence 11, Appl
11	26	74.3	487	4	US-09-352-159-21	Sequence 21, Appl
12	26	74.3	487	4	US-09-352-168-21	Sequence 21, Appl
13	26	74.3	554	4	US-09-352-159-17	Sequence 17, Appl
14	26	74.3	554	4	US-09-352-168-17	Sequence 17, Appl
15	26	74.3	591	4	US-09-352-159-44	Sequence 44, Appl
16	26	74.3	591	4	US-09-352-159-46	Sequence 46, Appl
17	26	74.3	598	4	US-09-352-159-40	Sequence 40, Appl
18	26	74.3	598	4	US-09-352-159-42	Sequence 42, Appl
19	26	74.3	600	4	US-09-352-159-23	Sequence 23, Appl
20	26	74.3	600	4	US-09-352-159-36	Sequence 36, Appl
21	26	74.3	600	4	US-09-352-159-38	Sequence 38, Appl
22	26	74.3	600	4	US-09-352-168-23	Sequence 23, Appl
23	26	74.3	692	4	US-09-352-159-19	Sequence 19, Appl
24	26	74.3	692	4	US-09-352-168-19	Sequence 19, Appl
25	26	74.3	759	2	US-08-637-759B-89	Sequence 89, Appl
26	26	74.3	759	3	US-08-871-355A-89	Sequence 89, Appl
27	26	74.3	829	4	US-09-352-159-33	Sequence 33, Appl

28	26	74.3	829	4	US-09-352-168-33	Sequence 33, Appl
29	26	74.3	991	4	US-09-352-159-27	Sequence 27, Appl
30	26	74.3	991	4	US-09-352-168-27	Sequence 27, Appl
31	26	74.3	1000	4	US-09-352-159-25	Sequence 25, Appl
32	26	74.3	1000	4	US-09-352-168-25	Sequence 25, Appl
33	26	74.3	1196	4	US-09-352-159-31	Sequence 31, Appl
34	26	74.3	1196	4	US-09-352-168-31	Sequence 31, Appl
35	26	74.3	1205	4	US-09-352-159-29	Sequence 29, Appl
36	26	74.3	1205	4	US-09-352-168-29	Sequence 29, Appl
37	25	71.4	58	3	US-08-405-647B-43	Sequence 43, Appl
38	25	71.4	58	4	US-08-985-499-43	Sequence 43, Appl
39	25	71.4	58	5	PCT-US96-03180-43	Sequence 43, Appl
40	25	71.4	159	4	US-09-518-046-24	Sequence 24, Appl
41	25	71.4	164	2	US-08-911-319A-1	Sequence 1, Appl
42	25	71.4	164	3	US-09-352-619-1	Sequence 1, Appl
43	25	71.4	283	3	US-08-807-151-1	Sequence 1, Appl
44	25	71.4	455	1	US-08-258-261B-18	Sequence 18, Appl
45	25	71.4	455	1	US-08-456-837-18	Sequence 18, Appl
46	25	71.4	455	1	US-08-457-342-18	Sequence 18, Appl
47	25	71.4	455	1	US-08-457-646A-18	Sequence 18, Appl
48	25	71.4	455	1	US-08-458-076A-18	Sequence 18, Appl
49	25	71.4	455	1	US-08-457-335A-18	Sequence 18, Appl
50	25	71.4	455	2	US-08-729-214-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1
US-08-608-241-2
; Sequence 2, Application US/08608241
; Patent No. 5747328
; GENERAL INFORMATION:
; APPLICANT: Donohue, Timothy J
; APPLICANT: Barber, Robert D
; APPLICANT: Withuhn, Vernon
; TITLE OF INVENTION: MICROBIAL SYSTEM FOR FORMALDEHYDE
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/608,241
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 960296.93511
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 376 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-608-241-2

Query Match 77.1%; Score 27; DB 1; Length 376;
Best Local Similarity 50.0%; Pred. No. 1.2e+02;

Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
1: |||
Db 316 WKGSARFG 323

RESULT 2
US-08-922-182-2
; Sequence 2, Application US/08922182
; Patent No. 5834300
; GENERAL INFORMATION:
; APPLICANT: Donohue, Timothy J
; APPLICANT: Barber, Robert D
; APPLICANT: Withuhn, Vernon
; TITLE OF INVENTION: MICROBIAL SYSTEM FOR FORMALDEHYDE
; TITLE OF INVENTION: SENSING AND REMEDIATION
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Plinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/922,182
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/608,241
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27,386
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 376 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-922-182-2

Query Match 77.1%; Score 27; DB 2; Length 376;
Best Local Similarity 50.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
1: |||
Db 316 WKGSARFG 323

RESULT 3
US-08-919-953-2
; Sequence 2, Application US/08919533
; Patent No. 5837481
; GENERAL INFORMATION:
; APPLICANT: Donohue, Timothy J
; APPLICANT: Barber, Robert D
; APPLICANT: Withuhn, Vernon
; TITLE OF INVENTION: MICROBIAL SYSTEM FOR FORMALDEHYDE
; TITLE OF INVENTION: SENSING AND REMEDIATION
; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Plinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/919,953
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/608,241
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27,386
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 376 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-919-953-2

Query Match 77.1%; Score 27; DB 2; Length 376;
Best Local Similarity 50.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
1: |||
Db 316 WKGSARFG 323

RESULT 4
US-09-192-983-2
; Sequence 2, Application US/09192983A
; Patent No. 6242244
; GENERAL INFORMATION:
; APPLICANT: Donohue, Timothy
; APPLICANT: Barber, Robert
; APPLICANT: Withuhn, Vernon
; TITLE OF INVENTION: Microbial System for Formaldehyde Sensing and
; TITLE OF INVENTION: Remediation
; FILE REFERENCE: 960296.95505
; CURRENT APPLICATION NUMBER: US/09/192,983A
; CURRENT FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: 08/919,953
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 08/608,241
; EARLIER FILING DATE: 1996-02-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Rhodobacter sphaeroides
US-09-192-983-2

Query Match 77.1%; Score 27; DB 4; Length 376;
Best Local Similarity 50.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
 11 11 1
 DB 316 WKGSAFCG 323

RESULT 5
 US-09-352-159-6
 ; Sequence 6, Application US/09352159A
 ; Patent No. 6211434
 ; GENERAL INFORMATION:
 ; APPLICANT: Duvick, Jonathan P.
 ; APPLICANT: Gilliam, Jacob T.
 ; APPLICANT: Maddox, Joyce R.
 ; TITLE OF INVENTION: Amino Polyol Amine Oxidase
 ; FILE REFERENCE: 1134
 ; CURRENT APPLICATION NUMBER: US/09/352,159A
 ; EARLIER FILING DATE: 1999-07-12
 ; EARLIER APPLICATION NUMBER: 60/092,936
 ; EARLIER FILING DATE: 1998-07-25
 ; EARLIER APPLICATION NUMBER: 60/135,391
 ; NUMBER OF SEQ ID NOS: 46
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 6
 ; LENGTH: 462
 ; TYPE: PRT
 ; ORGANISM: Exophiala spinifera
 US-09-352-159-6

Query Match 74.3%; Score 26; DB 4; Length 462;
 Best Local Similarity 50.0%; Pred. No. 2.4e+02;
 Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
 11 11 1
 DB 310 WREGGFSG 317

RESULT 6
 US-09-352-159-8
 ; Sequence 8, Application US/09352159A
 ; Patent No. 6211434
 ; GENERAL INFORMATION:
 ; APPLICANT: Duvick, Jonathan P.
 ; APPLICANT: Gilliam, Jacob T.
 ; APPLICANT: Maddox, Joyce R.
 ; TITLE OF INVENTION: Amino Polyol Amine Oxidase
 ; FILE REFERENCE: 1134
 ; CURRENT APPLICATION NUMBER: US/09/352,159A
 ; EARLIER FILING DATE: 1999-07-12
 ; EARLIER APPLICATION NUMBER: 60/092,936
 ; EARLIER FILING DATE: 1998-07-25
 ; EARLIER APPLICATION NUMBER: 60/135,391
 ; NUMBER OF SEQ ID NOS: 46
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 8
 ; LENGTH: 462
 ; TYPE: PRT
 ; ORGANISM: Exophiala spinifera
 US-09-352-159-8

Query Match 74.3%; Score 26; DB 4; Length 462;
 Best Local Similarity 50.0%; Pred. No. 2.4e+02;
 Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
 11 11 1

DB 310 WREGGFSG 317

RESULT 7
 US-09-352-168-6
 ; Sequence 6, Application US/09352168A
 ; Patent No. 6211435
 ; GENERAL INFORMATION:
 ; APPLICANT: Crasta, Oswald R.
 ; APPLICANT: Duvick, Jonathan P.
 ; APPLICANT: Folkerts, Otto
 ; APPLICANT: Gilliam, Jacob T.
 ; APPLICANT: Maddox, Joyce R.
 ; TITLE OF INVENTION: Amino Polyol Amine Oxidase
 ; FILE REFERENCE: 0875
 ; CURRENT APPLICATION NUMBER: US/09/352,168A
 ; EARLIER FILING DATE: 1999-07-12
 ; EARLIER APPLICATION NUMBER: 60/092,936
 ; EARLIER FILING DATE: 1998-07-25
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 6
 ; LENGTH: 462
 ; TYPE: PRT
 ; ORGANISM: Exophiala spinifera
 US-09-352-168-6

Query Match 74.3%; Score 26; DB 4; Length 462;
 Best Local Similarity 50.0%; Pred. No. 2.4e+02;
 Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
 11 11 1
 DB 310 WREGGFSG 317

RESULT 8
 US-09-352-168-8
 ; Sequence 8, Application US/09352168A
 ; Patent No. 6211435
 ; GENERAL INFORMATION:
 ; APPLICANT: Crasta, Oswald R.
 ; APPLICANT: Duvick, Jonathan P.
 ; APPLICANT: Folkerts, Otto
 ; APPLICANT: Gilliam, Jacob T.
 ; APPLICANT: Maddox, Joyce R.
 ; TITLE OF INVENTION: Amino Polyol Amine Oxidase
 ; FILE REFERENCE: 0875
 ; CURRENT APPLICATION NUMBER: US/09/352,168A
 ; EARLIER FILING DATE: 1999-07-12
 ; EARLIER APPLICATION NUMBER: 60/092,936
 ; EARLIER FILING DATE: 1998-07-25
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 8
 ; LENGTH: 462
 ; TYPE: PRT
 ; ORGANISM: Exophiala spinifera
 US-09-352-168-8

Query Match 74.3%; Score 26; DB 4; Length 462;
 Best Local Similarity 50.0%; Pred. No. 2.4e+02;
 Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
 11 11 1
 DB 310 WREGGFSG 317

```
RESULT 9
US-09-352-159-11
; Sequence 11, Application US/09352159A
; Patent No. 6211434
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jonathan P.
; APPLICANT: Gilliam, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; FILE REFERENCE: 1134
; CURRENT APPLICATION NUMBER: US/09/352,159A
; CURRENT FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; EARLIER APPLICATION NUMBER: 60/135,391
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-352-159-11
```

```
Query Match      74.3%; Score 26; DB 4; Length 463;
Best Local Similarity 50.0%; Pred. No. 2.4e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
OY 1 wrxxafxg 8
|| - ||
Db 311 WREGGFSG 318
```

```
RESULT 10
US-09-352-168-11
; Sequence 11, Application US/09352168A
; Patent No. 6211435
; GENERAL INFORMATION:
; APPLICANT: Crasta, Oswald R.
; APPLICANT: Duvick, Jonathan P.
; APPLICANT: Folkerts, Otto
; APPLICANT: Gilliam, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; FILE REFERENCE: 0875
; CURRENT APPLICATION NUMBER: US/09/352,168A
; CURRENT FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Exophiala spinifera
; FEATURE:
; OTHER INFORMATION: Extra lysine in the polypeptide sequence of
; OTHER INFORMATION: K:trpAPO, 463 aa.
US-09-352-168-11
```

```
Query Match      74.3%; Score 26; DB 4; Length 463;
Best Local Similarity 50.0%; Pred. No. 2.4e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
OY 1 wrxxafxg 8
|| - ||
Db 311 WREGGFSG 318
```

```
RESULT 11
US-09-352-159-21
; Sequence 21, Application US/09352159A
; Patent No. 6211434
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jonathan P.
; APPLICANT: Gilliam, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; FILE REFERENCE: 1134
; CURRENT APPLICATION NUMBER: US/09/352,159A
; CURRENT FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; EARLIER APPLICATION NUMBER: 60/135,391
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(24)
US-09-352-159-21
```

```
Query Match      74.3%; Score 26; DB 4; Length 487;
Best Local Similarity 50.0%; Pred. No. 2.6e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
OY 1 wrxxafxg 8
|| - ||
Db 335 WREGGFSG 342
```

```
RESULT 12
US-09-352-168-21
; Sequence 21, Application US/09352168A
; Patent No. 6211435
; GENERAL INFORMATION:
; APPLICANT: Crasta, Oswald R.
; APPLICANT: Duvick, Jonathan P.
; APPLICANT: Folkerts, Otto
; APPLICANT: Gilliam, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; FILE REFERENCE: 0875
; CURRENT APPLICATION NUMBER: US/09/352,168A
; CURRENT FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(24)
; OTHER INFORMATION: K:trpAPO translational fusion with barley alpha
; OTHER INFORMATION: amylase signal sequence, for expression and
; OTHER INFORMATION: secretion of the mature trpAPO in maize.
US-09-352-168-21
```

```
Query Match      74.3%; Score 26; DB 4; Length 487;
Best Local Similarity 50.0%; Pred. No. 2.6e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

OY 1 wrxxafxg 8
|| | |
Db 335 WREOGFSG 342

RESULT 13

US-09-352-159-17
; Sequence 17, Application US/09352159A
; Patent No. 6211434
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jonathan P.
; APPLICANT: Maddox, Joyce R.
; APPLICANT: Gilliam, Jacob T.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; TITLE OF INVENTION: Polynucleotides and Related Polypeptides and Methods of Use
; FILE REFERENCE: 1134
; CURRENT APPLICATION NUMBER: US/09/352,159A
; EARLIER FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; EARLIER APPLICATION NUMBER: 60/135,391
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 554
; TYPE: PRT
; ORGANISM: Exophiala spinfiera
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(89)
US-09-352-159-17

Query Match 74.3%; Score 26; DB 4; Length 554;
Best Local Similarity 50.0%; Pred. No. 2.9e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
|| | |
Db 402 WREOGFSG 409

RESULT 14

US-09-352-168-17
; Sequence 17, Application US/09352168A
; Patent No. 6211435
; GENERAL INFORMATION:
; APPLICANT: Crasla, Oswald R.
; APPLICANT: Duvick, Jonathan P.
; APPLICANT: Folkerts, Otto
; APPLICANT: Gilliam, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; TITLE OF INVENTION: Polynucleotides and Related Polypeptides and Methods of Use
; FILE REFERENCE: 0875
; CURRENT APPLICATION NUMBER: US/09/352,168A
; EARLIER FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 554
; TYPE: PRT
; ORGANISM: Exophiala spinfiera
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(89)
; OTHER INFORMATION: yeast alpha mating factor secretion signal.
US-09-352-168-17

Query Match 74.3%; Score 26; DB 4; Length 554;
Best Local Similarity 50.0%; Pred. No. 2.9e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
|| | |
Db 402 WREOGFSG 409

RESULT 15

US-09-352-159-44
; Sequence 44, Application US/09352159A
; Patent No. 6211434
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jonathan P.
; APPLICANT: Gilliam, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; TITLE OF INVENTION: Polynucleotides and Related Polypeptides and Methods of Use
; FILE REFERENCE: 1134
; CURRENT APPLICATION NUMBER: US/09/352,159A
; EARLIER FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; EARLIER APPLICATION NUMBER: 60/135,391
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 44
; LENGTH: 591
; TYPE: PRT
; ORGANISM: Rhinocladiaella atrovirens
US-09-352-159-44

Query Match 74.3%; Score 26; DB 4; Length 591;
Best Local Similarity 50.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 wrxxafxg 8
|| | |
Db 439 WREOGFSG 446

RESULT 16

US-09-352-159-46
; Sequence 46, Application US/09352159A
; Patent No. 6211434
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jonathan P.
; APPLICANT: Folkerts, Otto
; APPLICANT: Gilliam, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; TITLE OF INVENTION: Polynucleotides and Related Polypeptides and Methods of Use
; FILE REFERENCE: 1134
; CURRENT APPLICATION NUMBER: US/09/352,159A
; EARLIER FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; EARLIER APPLICATION NUMBER: 60/135,391
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 46
; LENGTH: 591
; TYPE: PRT
; ORGANISM: Rhinocladiaella atrovirens
US-09-352-159-46

Query Match 74.3%; Score 26; DB 4; Length 591;
Best Local Similarity 50.0%; Pred. No. 3.1e+02;

Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
|| | |
Db 442 WREGFGSG 449

RESULT 17
US-09-352-159-40
; Sequence 40, Application US/09352159A
; Patent No. 6211434
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jonathan P.
; APPLICANT: Gilliam, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; FILE REFERENCE: 1134
; CURRENT APPLICATION NUMBER: US/09/352,159A
; CURRENT FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; EARLIER APPLICATION NUMBER: 60/135,391
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 40
; LENGTH: 598
; TYPE: PRT
; ORGANISM: Exophiala spinifera
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (216)..(216)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-352-159-40

Query Match 74.3%; Score 26; DB 4; Length 598;
Best Local Similarity 50.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
|| | |
Db 446 WREGFGSG 453

RESULT 18
US-09-352-159-42
; Sequence 42, Application US/09352159A
; Patent No. 6211434
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jonathan P.
; APPLICANT: Gilliam, Jacob T.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; FILE REFERENCE: 1134
; CURRENT APPLICATION NUMBER: US/09/352,159A
; CURRENT FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; EARLIER APPLICATION NUMBER: 60/135,391
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 42
; LENGTH: 598
; TYPE: PRT
; ORGANISM: Rhinocladiaella atrovirens
US-09-352-159-42

Query Match 74.3%; Score 26; DB 4; Length 598;

Best Local Similarity 50.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
|| | |
Db 446 WREGFGSG 453

RESULT 19
US-09-352-159-23
; Sequence 23, Application US/09352159A
; Patent No. 6211434
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jonathan P.
; APPLICANT: Gilliam, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; FILE REFERENCE: 1134
; CURRENT APPLICATION NUMBER: US/09/352,159A
; CURRENT FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; EARLIER APPLICATION NUMBER: 60/135,391
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 23
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-352-159-23

Query Match 74.3%; Score 26; DB 4; Length 600;
Best Local Similarity 50.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 wrxxafxg 8
|| | |
Db 448 WREGFGSG 455

RESULT 20
US-09-352-159-36
; Sequence 36, Application US/09352159A
; Patent No. 6211434
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jonathan P.
; APPLICANT: Gilliam, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; FILE REFERENCE: 1134
; CURRENT APPLICATION NUMBER: US/09/352,159A
; CURRENT FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; EARLIER APPLICATION NUMBER: 60/135,391
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-352-159-36

Query Match 74.3%; Score 26; DB 4; Length 600;
Best Local Similarity 50.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 vxvxfxg 8
DB 448 WREGFGSG 455

Search completed: January 14, 2002, 07:23:45
Job time: 76 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:52:10 ; Search time 45.4 Seconds

(without alignments)
13.423 Million cell updates/sec

Title: 09-185908-1A

Sequence: 1 wxxxxsyxg 8

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 50 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25.0	4	2	A37832	phenol 2-monooxyge
2	25.0	5	2	I40469	dnazx-like protein
3	25.0	7	1	NVEG7	hypothalamic hepta
4	25.0	7	2	A60139	fatty-acid synthas
5	25.0	7	2	S33244	neuromodulatory pe
6	25.0	7	2	S33246	hypothetical L2 pr
7	25.0	7	4	I56695	glucose-6-phosphat
8	25.0	8	2	S11078	homeotic protein U
9	25.0	8	2	B27867	Ig heavy chain CRD
10	25.0	9	2	PT0324	cytokeratin 4 - bo
11	25.0	9	2	I46016	fructose-bisphosph
12	25.0	9	2	A28924	Ig heavy chain V r
13	25.0	9	2	S36850	T-cell receptor ga
14	25.0	9	2	I41946	gonadoliberin - pi
15	25.0	10	1	RHPCG	gonadoliberin I -
16	25.0	10	1	RHPCG	gonadoliberin - se
17	25.0	10	1	RHPCG	gonadoliberin - ch
18	25.0	10	1	RHPCG	gonadoliberin - m
19	25.0	10	1	RHPCG	gonadoliberin - p
20	25.0	10	1	RHPCG	gonadoliberin - s
21	25.0	10	2	A21114	gonadoliberin - se
22	25.0	10	2	PH0948	gonadoliberin - ch
23	25.0	11	2	A40693	transgelin - sheep
24	25.0	11	2	I60434	68kda neurofilamen
25	25.0	11	2	PH0941	T-cell receptor be
26	25.0	11	2	PH0938	T-cell receptor be
27	25.0	11	2	PH0914	T-cell receptor be
28	25.0	12	1	UOGM2	urotensin II - lon
29	25.0	12	2	A28856	fructose-bisphosph

30	25.0	12	2	S42765	urotensin II - tel
31	25.0	12	2	J50423	urotensin II-A pep
32	25.0	12	2	S26546	T-cell receptor be
33	25.0	12	2	S26556	T-cell receptor be
34	25.0	12	2	S26552	T-cell receptor be
35	25.0	12	2	S26556	T-cell receptor be
36	25.0	12	2	S26556	T-cell receptor be
37	25.0	12	2	S26556	T-cell receptor be
38	25.0	12	2	S26556	T-cell receptor be
39	25.0	12	2	S26556	T-cell receptor be
40	25.0	12	2	S26556	T-cell receptor be
41	25.0	12	2	S26556	T-cell receptor be
42	25.0	12	2	S26556	T-cell receptor be
43	25.0	12	2	S26556	T-cell receptor be
44	25.0	12	2	S26556	T-cell receptor be
45	25.0	12	2	S26556	T-cell receptor be
46	25.0	12	2	S26556	T-cell receptor be
47	25.0	12	2	S26556	T-cell receptor be
48	25.0	12	2	S26556	T-cell receptor be
49	25.0	12	2	S26556	T-cell receptor be
50	25.0	12	2	S26556	T-cell receptor be

ALIGNMENTS

RESULT 1
A37832
phenol 2-monooxygenase (EC 1.14.13.7) chain p5 - Pseudomonas sp. (strain CF600) (frag
C:Species: Pseudomonas sp.
C:Date: 14-Jun-1991 #sequence_revision 14-Jun-1991 #text_change 23-Jun-1993
R:Powlowski, J.; Shingler, V.
J. Bacteriol. 172, 6834-6840, 1990
A:Title: In vitro analysis of polypeptide requirements of multicomponent phenol hydro
A:Reference number: A37832; MUID:91072231
A:Accession: A37832
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-4 <POM>
C:Keywords: oxidoreductase

Query Match 25.0%; Score 2; DB 2; Length 4;
Best local similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 sy 6
DB 1 sy 2

RESULT 2
I40469
dnazx-like protein - Bacillus subtilis (fragment)
C:Species: Bacillus subtilis
C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 21-Jul-2000
R:Struck, J.C.; Hartmann, R.K.; Toschka, H.Y.; Erdmann, V.A.
Mol. Gen. Genet. 215, 478-482, 1989
A:Title: Transcription and processing of Bacillus subtilis small cytoplasmic RNA.
A:Reference number: I40469; MUID:89218958
A:Accession: I40469
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5 <RES>
A:Cross-references: EMBL:X14796; NID:g40130; PIDN:CAA32902.1; PID:g4376204
C:Genetics:
A:Start codon: GTG

Query Match 25.0%; Score 2; DB 2; Length 5;
Best local similarity 100.0%; Pred. No. 2.2e+05;

Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 sy 6
||
Db 2 sy 3

RESULT 3
NYPG7

hypothalamic heptapeptide - pig
C:Species: Sus scrofa domestica (domestic pig)
C>Date: 01-Sep-1981 #sequence_revision 01-Sep-1981 #text_change 23-Aug-1996
C:Accession: A01417
R:Chang, R.C.C.; Huang, W.Y.; Arimura, A.; Redding, T.W.; Coy, D.H.; Saifran, M.; Kong, Horm. Metab. Res. 13, 228-232, 1981
A>Title: Isolation, structure and synthesis of a heptapeptide with in vitro ACTH-release
A:Reference number: A01417; MUID:81213980
A:Accession: A01417
A:Molecule type: protein
A:Residues: 1-7 <CHAS>
C:Superfamily: hypothalamic heptapeptide
C:Keywords: hypothalamus

Query Match 25.0%; Score 2; DB 1; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sy 6
||
Db 5 sy 6

RESULT 4
A60139

fatty-acid synthase (EC 2.3.1.85) - rabbit (fragment)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C>Date: 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change 26-May-2000
C:Accession: A60139
R:Hardie, D.G.; Dewart, K.B.; Altken, A.; McCarthy, A.D. Biochim. Biophys. Acta 828, 380-382, 1985
A>Title: Amino acid sequence around the reactive serine residue of the thioesterase domain
A:Reference number: A60139; MUID:85175165
A:Accession: A60139
A:Molecule type: protein
A:Residues: 1-7 <HAK>
C:Superfamily: rat fatty-acid synthase; 3-oxoacyl-[acyl-carrier-protein] synthase I homology
C:Keywords: acyltransferase; carrier protein; coenzyme A; homodimer; multifunctional enzyme; Active site: Ser (of oleoyl-[acyl-carrier-protein] hydrolase) #status experimental

Query Match 25.0%; Score 2; DB 2; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sy 6
||
Db 5 sy 6

RESULT 5
S33244

neuromodulatory peptide Wwamide-1 - giant African snail
C:Species: Achatina fulica (giant African snail)
C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Jul-1997
C:Accession: S33244
R:Minakata, H.; Ikeda, T.; Muneko, Y.; Kobayashi, M.; Nomoto, K. FEBS Lett. 323, 104-108, 1993
A>Title: Wwamide-1, -2 and -3, novel neuromodulatory peptides isolated from ganglia of t
A:Reference number: S33244; MUID:93265912
A:Accession: S33244
A:Status: preliminary

A:Molecule type: protein
A:Residues: 1-7 <MIN>

Query Match 25.0%; Score 2; DB 2; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wk 2
||
Db 1 wk 2

RESULT 6
S33246

neuromodulatory peptide Wwamide-3 - giant African snail
C:Species: Achatina fulica (giant African snail)
C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Jul-1997
C:Accession: S33246
R:Minakata, H.; Ikeda, T.; Muneko, Y.; Kobayashi, M.; Nomoto, K. FEBS Lett. 323, 104-108, 1993
A>Title: Wwamide-1, -2 and -3, novel neuromodulatory peptides isolated from ganglia o
A:Reference number: S33244; MUID:93265912
A:Accession: S33246
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-7 <MIN>

Query Match 25.0%; Score 2; DB 2; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wk 2
||
Db 1 wk 2

RESULT 7
J56695

hypothetical L2 protein (mistranslated) - human papillomavirus type 16 (fragment)
C:Species: human papillomavirus type 16
C>Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 18-Feb-2000
C:Accession: J56695
R:Schneider-Maunoury, S.; Croissant, O.; Orth, G. J. Virol. 61, 3295-3298, 1987
A>Title: Integration of human papillomavirus type 16 DNA sequences: a possible early
A:Reference number: J56695; MUID:87311896
A:Accession: J56695
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-7 <SCH>
A:Cross-references: GB:M30709; NID:g190253; PIDN:AAA65995.1; PTD:g553616
C:Comment: This is the hypothetical translation of a viral sequence integrated into t
C:Comment: It is translated in an incorrect, -1, reading frame of the L2 protein.

Query Match 25.0%; Score 2; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sy 6
||
Db 2 sy 3

RESULT 8
S11078

glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) - yeast (Pichia jadinii) (fragment)
C:Species: Pichia jadinii, Candida utilis
C>Date: 30-Jun-1991 #sequence_revision 30-Sep-1991 #text_change 05-Aug-1994
C:Accession: S11078
R:Egestad, B.; Estonius, M.; Danielsson, O.; Persson, B.; Cederlund, E.; Kaiser, R.;

FEBS Lett. 269, 194-196, 1990
A:Title: Fast atom bombardment mass spectrometry and chemical analysis in determinations
A:Reference number: S11074; MUID:90353571
A:Accession: S11078

A:Molecule type: protein
A:Residues: 1-8 <EGE>

A>Note: the source is designated as Pichia jadinii

C:Keywords: acetylated amino end; oxidoreductase; pentose phosphate pathway
P1/Modified site: acetylated amino end (Ser) #status experimental

Query Match
Best Local Similarity 100.0%; Score 2; DB 2; Length 8;
Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
||
Db 1 sy 2

RESULT 9

B27867
homeotic protein Ultrabithorax - fruit fly (Drosophila melanogaster) (fragment)
C:Species: Drosophila melanogaster
C:Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 23-Feb-1997
C:Accession: B27867

R:SaarL, G.; Bienz, M.
EMBO J. 6, 1775-1779, 1987

A:Title: The structure of the ultrabithorax promoter of Drosophila melanogaster.
A:Reference number: A91072
A:Accession: B27867

A:Molecule type: mRNA
A:Residues: 1-8 <SNA>

C:Genetics:

A:Gene: FlyBase:UBX

A:Cross-references: FlyBase:FBgn0003944

C:Keywords: DNA binding; nucleus; transcription regulation

Query Match
Best Local Similarity 100.0%; Score 2; DB 2; Length 8;
Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
||
Db 3 sy 4

RESULT 10

PT0238
Ig heavy chain CRD3 region (clone 2-94B) - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Aug-1996
C:Accession: PT0238

R:Yamada, M.; Wasserman, R.; Reichard, B.A.; Shane, S.; Caton, A.J.; Rovera, G.
J. Exp. Med. 173, 395-407, 1991

A:Title: Preferential utilization of specific immunoglobulin heavy chain diversity and J

A:Reference number: PT0222; MUID:91108337

A:Accession: PT0238

A:Molecule type: DNA

A:Residues: 1-9 <YAM>

A:Experimental source: B lymphocyte

C:Keywords: heterotetramer; immunoglobulin

Query Match
Best Local Similarity 100.0%; Score 2; DB 2; Length 9;
Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
||
Db 7 sy 8

RESULT 11
PT0324
Ig heavy chain CRD3 region (clone J2-106C) - human (fragment)

C:Species: Homo sapiens (man)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Aug-1996
C:Accession: PT0324

R:Yamada, M.; Wasserman, R.; Reichard, B.A.; Shane, S.; Caton, A.J.; Rovera, G.
J. Exp. Med. 173, 395-407, 1991

A:Title: Preferential utilization of specific immunoglobulin heavy chain diversity an

A:Reference number: PT0222; MUID:91108337

A:Accession: PT0324

A:Molecule type: DNA

A:Residues: 1-9 <YAM>

A:Experimental source: B lymphocyte

C:Keywords: heterotetramer; immunoglobulin

Query Match
Best Local Similarity 100.0%; Score 2; DB 2; Length 9;
Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
||
Db 7 sy 8

RESULT 12

I46016

cytokeratin 4 - bovine (fragment)

C:Species: Bos primigenius taurus (cattle)

C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 05-Nov-1999
C:Accession: I46016

R:Blessing, M.; Jorcano, J.L.; Franke, W.W.
EMBO J. 8, 117-126, 1989

A:Title: Enhancer elements directing cell-type-specific expression of cytokeratin gen

A:Reference number: I46016; MUID:89231609

A:Accession: I46016

A>Status: preliminary; translated from GR/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-9 <BLR>

A:Cross-references: EMBL:X14478; NID:9303; PIDN:CMA32640.1; PID:9577897

Query Match
Best Local Similarity 100.0%; Score 2; DB 2; Length 9;
Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
||
Db 2 sy 3

RESULT 13

A28924

fructose-bisphosphate aldolase (EC 4.1.2.13) B, hepatic - rabbit (fragment)

C:Species: Oryctolagus cuniculus (domestic rabbit)

C:Date: 28-Aug-1989 #sequence_revision 28-Aug-1989 #text_change 07-Feb-1997
C:Accession: A28924

R:Lacko, A.G.; Brox, L.W.; Gracy, R.W.; Horecker, B.L.
J. Biol. Chem. 245, 2140-2141, 1970

A:Title: The carboxyl-terminal structure of rabbit liver aldolase (aldolase B).

A:Reference number: A28924; MUID:70166720

A:Accession: A28924

A:Molecule type: protein

A:Residues: 1-9 <LAK>

C:Keywords: aldehyde-lyase; carbon-carbon lyase; gluconeogenesis; glycolysis; liver;

Query Match
Best Local Similarity 100.0%; Score 2; DB 2; Length 9;
Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 SY 6
||
DB 8 SY 9

RESULT 14

S36850
Ig heavy chain V region - mouse
C:Species: Mus musculus (house mouse)
C:Date: 06-Jan-1994 #sequence_revision 01-Dec-1995 #text_change 05-Nov-1999
C:Accession: S36850
R:Jacob, J.; Kelsoe, G.
Submitted to the EMBL Data Library, July 1992
A:Description: In situ studies on the primary immune response to (4-hydroxy-3-nitrophenyl
A:Reference number: S25024
A:Accession: S36850
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-9 <JAC>
A:Cross-references: EMBL:X67387; NID:950113; PIDN:CAA47799.1; PID:e51594; PID:g1333871
C:Keywords: heterotrimer; immunoglobulin

Query Match 25.0%; Score 2; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 SY 6
||
DB 7 SY 8

RESULT 15

G41946
T-cell receptor gamma chain (2t.23) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 07-May-1999
C:Accession: G41946
R:Whitsett, M.; Mosley, R.L.; Whitsett, L.; Schaefer, F.V.; Miller, K.S.; Klein, J.R.
Mol. Cell. Biol. 11, 5902-5909, 1991
A:Title: Rearrangement and junctional-site sequence analyses of T-cell receptor gamma ge
A:Reference number: A41946; MUID:92049316
A:Accession: G41946
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-9 <WHE>
C:Keywords: T-cell receptor

Query Match 25.0%; Score 2; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 SY 6
||
DB 1 SY 2

RESULT 16

RHPG3
gonadoliberin - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 18-Mar-1997
C:Accession: A01411
R:Baba, Y.; Matsuo, H.; Schally, A.V.
Biochem. Biophys. Res. Commun. 44, 459-463, 1971
A:Title: Structure of the porcine LH- and FSH-releasing hormone. II. Confirmation of the
A:Reference number: A90172; MUID:72114303
A:Accession: A01411
A:Molecule type: protein
A:Residues: 1-10 <BAB>
R:Matsuo, H.; Arimura, A.; Nair, R.M.G.; Schally, A.V.
Biochem. Biophys. Res. Commun. 45, 822-827, 1971

A:Title: Synthesis of the porcine LH- and FSH-releasing hormone by the solid-phase me
A:Reference number: A90176; MUID:72065376

A:Contents: annotation; synthesis
A:Note: the synthetic and natural hormones have the same physicochemical and biologic
R:Baba, Y.; Arimura, A.; Schally, A.V.
Biochem. Biophys. Res. Commun. 45, 483-487, 1971
A:Title: On the tryptophan residue in porcine LH and FSH-releasing hormone.
A:Reference number: A90175; MUID:72117544
A:Contents: annotation
A:Note: Trp-3 appears to be essential for biological activity
C:Comment: This hypothalamic hormone stimulates the secretion of both luteinizing and
C:Superfamily: gonadoliberin
C:Keywords: amidated carboxyl end; hormone; hypothalamus; pyroglutamic acid
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F:10/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 25.0%; Score 2; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 SY 6
||
DB 4 SY 5

RESULT 17

RHSHG
gonadoliberin - sheep
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 18-Mar-1997
C:Accession: A93780; A01411
R:Burgus, R.; Butcher, M.; Amoss, M.; Ling, N.; Monahan, M.; Rivier, J.; Fellows, R.;
Proc. Natl. Acad. Sci. U.S.A. 69, 278-282, 1972
A:Title: Primary structure of the ovine hypothalamic luteinizing hormone-releasing fa
A:Reference number: A93780; MUID:72094314
A:Accession: A93780
A:Molecule type: protein
A:Residues: 1-10 <BUR>
A:Note: the natural and synthetic hormones have the same biological activity
C:Comment: This hypothalamic hormone stimulates the secretion of both luteinizing and
C:Superfamily: gonadoliberin
C:Keywords: amidated carboxyl end; hormone; hypothalamus; pyroglutamic acid
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F:10/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 25.0%; Score 2; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 SY 6
||
DB 4 SY 5

RESULT 18

RHAQ1
gonadoliberin I - American alligator
N:Alternate names: gonadotropin-releasing hormone I
C:Species: Alligator mississippiensis (American alligator)
C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 18-Mar-1997
C:Accession: A60066
R:Lovejoy, D.A.; Fischer, W.H.; Parker, D.B.; McRoy, J.E.; Park, M.; Lance, V.; Swan
Regul. Pept. 33, 105-116, 1991
A:Title: Primary structure of two forms of gonadotropin-releasing hormone from brains
A:Reference number: A60066; MUID:91352338
A:Accession: A60066
A:Molecule type: protein
A:Residues: 1-10 <LOV>
C:Superfamily: gonadoliberin
C:Keywords: amidated carboxyl end; hormone; hypothalamus; pyroglutamic acid
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

F:10/Modified site: amidated carboxyl end (gly) #status experimental

Query Match
Best Local Similarity 100.0%; Score 2; DB 1; Length 10;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
||
Db 4 sy 5

RESULT 19

RILMGS
gonadoliberin - sea lamprey
N:Alternate names: gonadotropin releasing hormone (GNRH)
C:Species: Petromyzon marinus (sea lamprey)
C>Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 18-Mar-1997
C:Accession: A01412
R:Sherwood, N.M.; Sower, S.A.; Marshak, D.R.; Fraser, B.A.; Brownstein, M.J.
J. Biol. Chem. 261, 4812-4819, 1986
A:Title: Primary structure of gonadotropin-releasing hormone from lamprey brain.
A:Reference number: A01412; MUID:86168192
A:Accession: A01412
A:Molecule type: protein
A:Residues: 1-10 <SHE>
C:Comment: This hormone was isolated from the brain.

C:Superfamily: gonadoliberin
F:1/Modified site: amidated carboxyl end; hormone: pyroglutamic acid
F:2/Modified site: pyroglutamate carboxylic acid (Gln) #status experimental
F:10/Modified site: amidated carboxyl end (gly) #status experimental

Query Match
Best Local Similarity 100.0%; Score 2; DB 1; Length 10;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
||
Db 7 wk 8

RESULT 20

A49187
gonadotropin-releasing hormone III - sea lamprey
C:Species: Petromyzon marinus (sea lamprey)
C>Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 03-Mar-1995
C:Accession: A49187
R:Sower, S.A.; Chiang, Y.C.; Lovas, S.; Conlon, J.M.
Endocrinology 132, 1125-1131, 1993
A:Title: Primary structure and biological activity of a third gonadotropin-releasing hor

A:Reference number: A49187; MUID:93178316
A:Accession: A49187
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-10 <SOM>
A:Experimental source: brain
A:Note: sequence extracted from NCBI backbone (NCBIP:126381)

Query Match
Best Local Similarity 100.0%; Score 2; DB 2; Length 10;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
||
Db 7 wk 8

RESULT 21

A21114
gonadoliberin - chum salmon
C:Species: Oncorhynchus keta (chum salmon)

C>Date: 10-Aug-1990 #sequence_revision 10-Aug-1990 #text_change 18-Jun-1993
C:Accession: A21114
R:Sherwood, N.; Elden, L.; Brownstein, M.; Spiess, J.; Rivier, J.; Vale, W.
Proc. Natl. Acad. Sci. U.S.A. 80, 2794-2798, 1983
A:Title: Characterization of a teleost gonadotropin-releasing hormone.
A:Reference number: A21114; MUID:83195140
A:Accession: A21114
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-10 <SHE>

Query Match
Best Local Similarity 100.0%; Score 2; DB 2; Length 10;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
||
Db 4 sy 5

RESULT 22

PH0948
T-cell receptor beta chain V-D-J region - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C>Date: 09-Oct-1992 #sequence_revision 19-Oct-1995 #text_change 30-May-1997
C:Accession: PH0948; PH0897; PH0909; PH0899
R:Gold, D.P.; Offner, H.; Sun, D.; Wiley, S.; Vandenbark, A.A.; Wilson, D.B.
J. Exp. Med. 174, 1467-1476, 1991
A:Title: Analysis of T cell receptor beta chains in Lewis rats with experimental alle
A:Reference number: PH0891; MUID:92078857
A:Accession: PH0948
A:Molecule type: mRNA
A:Residues: 1-10 <G01>
A:Experimental source: myelin basic protein fragment-reactive T-cell, recovered from

A:Accession: PH0897
A:Molecule type: mRNA
A:Residues: 1-10 <G02>
A:Experimental source: myelin basic protein-immunized T-cell, clones 3, 6-2, 14, hybr

A:Accession: PH0909
A:Molecule type: mRNA
A:Residues: 1-10 <G03>

A:Accession: PH0899
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-10 <G04>

A:Experimental source: clone 14
C:Keywords: T-cell receptor

Query Match
Best Local Similarity 100.0%; Score 2; DB 2; Length 10;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
||
Db 7 sy 8

RESULT 23

A40693
transgelin - sheep (fragment)
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C>Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 31-Oct-1997
C:Accession: A40693
R:Shapland, C.; Hsuan, J.J.; Totty, N.F.; Lawson, D.
J. Cell Biol. 121, 1065-1073, 1993
A:Title: Purification and properties of transgelin: a transformation and shape change

A:Reference number: A40693; MUID:93273790
A:Accession: A40693
A:Molecule type: protein
A:Residues: 1-11 <SHA>
A:Experimental source: aorta

C;Comment: This protein gels actin and is down regulated by transformation or loss of cell
C;Superfamily: smooth muscle protein SM22; calponin repeat homology; smooth muscle protein
C;Keywords: actin binding; cytoskeleton

Query Match 25.0%; Score 2; DB 2; Length 11;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 SY 6
II
Db 4 SY 5

RESULT 24
I60434
68kDa neurofilament - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999
C;Accession: I60434
R;Reeбен, M.; Neuman, T.; Palgi, J.; Palm, K.; Paalme, V.; Saarma, M.
J. Neurosci. Res. 40, 177-188, 1995
A;Title: Characterization of the rat light neurofilament (NF-L) gene promoter and identification
A;Reference number: I60434; MUID:95264348
A;Accession: I60434
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-11 <RES>
A;Cross-references: EMBL:X53981; NID:9452676; PIDD:CNA37931.1; PID:9452677
C;Genetics:
A;Gene: NF68

Query Match 25.0%; Score 2; DB 2; Length 11;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 SY 6
II
Db 5 SY 6

RESULT 25
PH0941
T-cell receptor beta chain V-D-J region (clone 12) - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 09-Oct-1992 #sequence_revision 09-Oct-1992 #text_change 30-May-1997
C;Accession: PH0941
R;Gold, D.P.; Offner, H.; Sun, D.; Wiley, S.; Vandenbark, A.A.; Wilson, D.B.
J. Exp. Med. 174, 1467-1476, 1991
A;Title: Analysis of T cell receptor beta chains in Lewis rats with experimental allergy
A;Reference number: PH0941; MUID:92078857
A;Accession: PH0941
A;Molecule type: mRNA
A;Residues: 1-11 <GOL>
A;Experimental source: complete Freund's adjuvant-immunized lymph node
C;Keywords: T-cell receptor

Query Match 25.0%; Score 2; DB 2; Length 11;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 SY 6
II
Db 8 SY 9

Search completed: January 14, 2002, 07:58:36
Job time: 386 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:56:35 ; Search time 30.66 seconds
(without alignments)
9.567 Million cell updates/sec

Title: 09-185908-1A

Perfect score: 8

Sequence: 1 wxxsxyxg 8

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size : 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 50 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25.0	7	1	HY7_PIG	P01153 sus scrofa
2	25.0	7	1	WMA2_ACHFV	P35920 achalina fu
3	25.0	7	1	WMA2_ACHFV	P35921 achalina fu
4	25.0	8	1	AL16_CARMA	P81819 carcinus ma
5	25.0	10	1	GON1_ALLMT	P37041 alligator m
6	25.0	10	1	GON1_PETMA	P20367 petromyzon
7	25.0	10	1	GON3_ONCKE	P30948 oncorhynch
8	25.0	10	1	RLA2_MOUSE	P99027 mus muscul
9	25.0	12	1	UR2A_CATCO	P04558 catostomus
10	25.0	12	1	UR2B_CATCO	P04559 catostomus
11	25.0	12	1	UR2B_CYPCA	P04561 cyprinus ca
12	25.0	12	1	UR2_GILMT	P01147 gillichthys
13	25.0	12	1	UR2_POLSP	P81022 polyodon sp
14	25.0	12	1	UR2_SCYCA	P35490 scyllorhinu
15	25.0	13	1	MLA_ANOCA	P41589 anolis caro
16	25.0	13	1	MLA_CANCR	P01198 canelis dfo
17	25.0	14	1	MAST_PARI0	P42716 parapolybia
18	25.0	14	1	MAST_POLJA	P01517 polistes ja
19	25.0	14	1	MAST_VESXA	P01515 vespa xanth
20	25.0	14	1	SMS_MYOSC	P20750 myoxcephal
21	25.0	14	1	SMS_ALMT	P31885 alligator m
22	25.0	15	1	CMX2_CONAL	P56640 conus aulic
23	25.0	15	1	MAUT_BACTO	P80072 bacillus th
24	25.0	15	1	PH2_PERAM	P86595 petriplaneta
25	25.0	16	1	CMX1_CONAL	P56639 conus aulic
26	25.0	16	1	CMX1_CONAL	P56641 conus aulic
27	25.0	16	1	RIPK_TRIKI	P16093 trichosantha
28	25.0	17	1	PH3_PERAM	P82696 petriplaneta
29	25.0	17	1	TPIS_PINPS	P81666 pinus pinas
30	25.0	18	1	AG1_EUPCH	P33888 euphorbia c
31	25.0	18	1	AL12_CVDPO	P81253 cydia pomon
32	25.0	18	1	FMF1_ECOLI	P20860 escherichia
33	25.0	18	1		

34	2	25.0	18	1	PHPT_PSESE	P25271 pseudaleria
35	2	25.0	19	1	COXR_THUOB	P80984 thunnus obe
36	2	25.0	19	1	DHAB_COMTE	P80704 comamonas t
37	2	25.0	19	1	DHAB_COMTE	P86503 streptomyc
38	2	25.0	20	1	CAT4_FASHE	P80528 fasciola he
39	2	25.0	20	1	CISY_STRHY	P20903 streptomyc
40	2	25.0	20	1	CISY_STRHY	P19094 mustelus ca
41	2	25.0	20	1	LPJ3_HUMAN	P56643 homo sapien
42	2	25.0	20	1	LPJ3_HUMAN	P37362 pyrrhocoris
43	2	25.0	20	1	PIRX_CUCPE	P80750 cucurbita p
44	2	25.0	20	1	RLC1_HALMA	P12740 halocaula p
45	2	25.0	20	1	VR90_BORPE	P81549 bordelella
46	2	25.0	21	1	FIBB_ANTAM	P14465 antilocapra
47	2	25.0	21	1	NRLA_ACISP	P33036 actinobact
48	2	25.0	21	1	PSRF_SYNVU	P12239 synechococ
49	2	25.0	21	1	TEPT_ATIME	P56587 apis mellif
50	2	25.0	22	1	LANN_STRMU	P80666 streptococc

ALIGNMENTS

RESULT 1
ID HY7_PIG STANDARD: PRT: 7 AA.
AC P01153;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 21-JUL-1986 (Rel. 01, Last annotation update)
DE HYPOTHALMIC HEPYAPEPTIDE.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
OX NCBI_TaxID:9623;
RN [1]
RP SEQUENCE, AND SYNTHESIS.
RX MEDLINE=81213980; PubMed=6263778;
RA Chang R.C.C., Huang W.-Y., Arimura A., Redding T.W., Coy D.H.,
RA Saffran M., Kong A., Hamilton J.W., Cohn D.V., Schally A.V.;
RT "Isolation, structure and synthesis of a heptapeptide with in vitro
RT ACPH-releasing activity from porcine hypothalamus.";
DR PIR: A01417; NRP67.
SQ SEQUENCE 7 AA: 957 MW: 632B45B1FB5059A0 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 7;
Best Local Similarity 100.0%; Pred. No. 1e+05; 0; Indels 0; Gaps 0;
Matches 2; Conservative 0; Mismatches 0;

QY 5 sy 6
Db 5 sy 6
RESULT 2
ID WMA2_ACHFV STANDARD: PRT: 7 AA.
AC P35920;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE WMAWIDE-2.
OS Achalina fulica (Giant African snail).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;
OC Achatinacea; Achatinidae; Achalina.
OX NCBI_TaxID=6530;
RN [1]
RP SEQUENCE.
RC TISSUE=Ganglion;
RC MEDLINE=93265912; PubMed=8495720;
RA Minakata H., Ikeda T., Muneoka Y., Kobayashi M., Nomoto K.;
RT "Wamide-1, -2 and -3: novel neuromodulatory peptides isolated from

RT ganglia of the African giant snail, Achatina fulica.";
 RL FEBS Lett. 323:104-108(1993).
 KW Neuropeptide; Amidation.
 FT MOD_RES 7
 SO SEQUENCE 7 AA; 964 MW; 7362D5B686D32310 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 7;
 Best Local Similarity 100.0%; Pred. No. 1e+05;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
 Db 1 WK 2

RESULT 3
 WMA3_ACHFU STANDARD; PRT; 7 AA.
 ID WMA3_ACHFU
 AC P35921;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE WMA3_ACHFU-3
 OS Achatina fulica (Giant African snail).
 OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;
 OC Achatinacea; Achatinidae; Achatina.
 OX NCBI_TaxID=6530;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Ganglion;
 RX MEDLINE=93265912; PubMed=8495720;
 RA Minakata H., Ikeda T., Muneoka Y., Kobayashi M., Nomoto K.;
 RT "Wamide-1, -2 and -3: novel neuromodulatory peptides isolated from
 ganglia of the African giant snail, Achatina fulica.";
 RL FEBS Lett. 323:104-108(1993).
 RL PIR: S33244; S33244.
 KW Neuropeptide; Amidation.
 FT MOD_RES 7
 SO SEQUENCE 7 AA; 965 MW; 7362D5B69B132310 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 7;
 Best Local Similarity 100.0%; Pred. No. 1e+05;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
 Db 1 WK 2

RESULT 4
 AL16_CARMA STANDARD; PRT; 8 AA.
 ID AL16_CARMA
 AC P81819;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE CARCINUSTATIN 16.
 OS Carcinus maenas (Common shore crab) (Green crab).
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
 OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
 OC Eubrachyura; Portunoidae; Portunidae; Carcinus.
 OX NCBI_TaxID=6759;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Cerebral ganglion, and thoracic ganglion;
 RX MEDLINE=98121193; PubMed=9461295;
 RA Thøgers A., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
 RT Isolation and identification of multiple neuropeptides of the
 allatostatin superfamily in the shore crab Carcinus maenas.";
 RL Eur. J. Biochem. 250:727-734(1997).

CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
 CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
 CC Neuropeptide; Amidation; Multigene family.
 KW MOD_RES 8
 FT MOD_RES 8
 SO SEQUENCE 8 AA; 813 MW; 7C286B45AB476878 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 8;
 Best Local Similarity 100.0%; Pred. No. 1e+05;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
 Db 5 SY 6

RESULT 5
 GON1_ALIMI STANDARD; PRT; 10 AA.
 ID GON1_ALIMI
 AC P37041; P20407;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE GONADOLIBERIN I (GONADOTROPIN-RELEASING HORMONE I) (GNRH-I) (LH-RH I)
 DE (LH-RH I).
 OS Alligator mississippiensis (American alligator).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Crocodylidae; Alligatorineae; Alligator.
 OX NCBI_TaxID=8496;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Brain;
 RX MEDLINE=91352338; PubMed=1882082;
 RA Lovejoy D.A., Fischer W.H., Parker D.B., McRory J.E., Park M.,
 RA Lance V., Swanson P., Rivier J.E., Sherwood N.M.;
 RT "Primary structure of two forms of gonadotropin-releasing hormone
 from brains of the American alligator (Alligator mississippiensis).";
 RL Regul. Pept. 33:105-116(1991).
 RL -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS.
 CC -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.
 CC PIR: A60066; RHAOL
 DR InterPro: IPR002012; GNRH.
 DR Pfam: PF00446; GNRH; 1.
 DR PROSITE: PS00473; GNRH; 1.
 KW Hormone; Amidation; Hypothalamus.
 FT MOD_RES 1
 FT MOD_RES 10
 SO SEQUENCE 10 AA; 1172 MW; 284B23D7286B45A3 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 10;
 Best Local Similarity 100.0%; Pred. No. 1e+05;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
 Db 4 SY 5

RESULT 6
 GON1_PETMA STANDARD; PRT; 10 AA.
 ID GON1_PETMA
 AC P04378;
 DT 20-MAR-1987 (Rel. 04, Created)
 DT 20-MAR-1987 (Rel. 04, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE GONADOLIBERIN I (GONADOTROPIN-RELEASING HORMONE I) (GNRH-I)
 DE (LH-RH I).
 OS Petromyzon marinus (Sea lamprey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
 OC Petromyzontiformes; Petromyzontidae; Petromyzon.
 OX NCBI_TaxID=7757;
 RN [1]

RP SEQUENCE.
 RC TISSUE=Brain;
 RX MEDLINE=86168192; PubMed=3514603;
 RA Sherwood N.M., Sower S.A., Marshak D.R., Fraser B.A., Brownstein M.J.;
 RT "Primary structure of gonadotropin-releasing hormone from lamprey brain."
 RL J. Biol. Chem. 261:4812-4819(1986).
 CC -1- FUNCTION: STIMULATES THE SECRETION OF BOTH LUTEINIZING AND
 FOLLICLE-STIMULATING HORMONES.
 CC -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.
 DR PIR: A01412; RHLMS.
 DR InterPro: IPR002012; GNRH.
 DR Pfam: PF00446; GNRH; 1.
 DR PROSITE: PS00473; GNRH; 1.
 KW Hormone; Amidation; Hypothalamus.
 FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
 FT MOD_RES 10 10 AMIDATION.
 SQ SEQUENCE 10 AA: 1244 MW: 1E4B36237B1735AB CRC64:

Query Match 25.0%; Score 2; DB 1; Length 10;
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
 11
 DB 7 WK 8

RESULT 7
 CON3_ONCKE
 ID CON3_ONCKE STANDARD: PRT: 10 AA.
 AC P20367; P81751;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE GONADOLIBERIN III (GONADOTROPIN-RELEASING HORMONE III) (GNRH-III) (LH-RH III) (LULIBERIN III).
 GN CNRH3.
 OS Oncorhynchus keta (Chum salmon), and
 OS Clupea pallasii (Pacific herring).
 OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
 OX NCBI_TaxID=8018; 30724;
 RN [1]
 RP SEQUENCE.
 RC SPECIES=O.keta;
 RX MEDLINE=83195140; PubMed=6341999;
 RA Sherwood N., Elden L., Brownstein M., Splies J., Rivier J., Vale W.;
 RT "Characterization of a teleost gonadotropin-releasing hormone."
 RL Proc. Natl. Acad. Sci. U.S.A. 80:2794-2798(1983).
 RN [2]
 RP SEQUENCE, AND FUNCTION.
 RC SPECIES=C.pallasii; TISSUE=Brain, and Pituitary;
 RX MEDLINE=20114351; PubMed=10650929;
 RA Carolisfeld J., Powell J.F.F., Park M., Fischer W.H., Craig A.G.,
 Chang J.P., Rivier J.E., Sherwood N.M.;
 RT "Primary structure and function of three gonadotropin-releasing hormones, including a novel form, from an ancient teleost, herring."
 RL Endocrinology 141:505-512(2000).
 CC -1- FUNCTION: STIMULATES THE SECRETION OF BOTH LUTEINIZING AND FOLLICLE-STIMULATING HORMONES.
 CC -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.
 DR PIR: A21114; A21114.
 DR InterPro: IPR002012; GNRH.
 DR Pfam: PF00446; GNRH; 1.
 DR PROSITE: PS00473; GNRH; 1.
 KW Hormone; Amidation; Hypothalamus.
 FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
 FT MOD_RES 10 10 AMIDATION.
 SQ SEQUENCE 10 AA: 1230 MW: 284B32337B6B45A3 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 10;
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
 11
 DB 4 sy 5

RESULT 8
 CON3_PETMA
 ID CON3_PETMA STANDARD: PRT: 10 AA.
 AC P30948;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE GONADOLIBERIN III (GONADOTROPIN-RELEASING HORMONE III) (GNRH-III) (LULIBERIN III).
 OS Petromyzon marinus (Sea lamprey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
 OC Petromyzontiformes; Petromyzontidae; Petromyzon.
 OX NCBI_TaxID=7757;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Brain;
 RX MEDLINE=93178316; PubMed=8440174;
 RA Sower S.A., Chiang Y.-C., Lovas S., Conlon J.M.;
 RT "Primary structure and biological activity of a third gonadotropin-releasing hormone from lamprey brain."
 RL Endocrinology 132:1125-1131(1993).
 CC -1- FUNCTION: STIMULATES THE SECRETION OF BOTH LUTEINIZING AND FOLLICLE-STIMULATING HORMONES.
 CC -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.
 DR InterPro: IPR002012; GNRH.
 DR Pfam: PF00446; GNRH; 1.
 DR PROSITE: PS00473; GNRH; 1.
 KW Hormone; Amidation; Hypothalamus.
 FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
 FT MOD_RES 10 10 AMIDATION.
 SQ SEQUENCE 10 AA: 1277 MW: 284B36237A1A1F5A3 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 10;
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
 11
 DB 7 WK 8

RESULT 9
 RLA2_MOUSE
 ID RLA2_MOUSE STANDARD: PRT: 10 AA.
 AC P99027;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE 60S ACIDIC RIBOSOMAL PROTEIN P2 (FRAGMENT).
 GN RPLP2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Liver;
 RA Sanchez J.-C., Rouge V., Frutiger S., Hughes G., Yan J.X.,
 Hoogland C., Appel R.D., Binz P.-A., Hochstrasser D.F.,
 Cowthorne M.;
 RT submitted (AUG-1998) to the SWISS-PROV data bank.
 RL -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE ELONGATION STEP OF

```

CC PROTEIN SYNTHESIS.
CC -I SUBUNIT: P1 AND P2 EXIST AS DIMERS AT THE LARGE RIBOSOMAL
CC SUBUNIT.
CC -I PTM: PHOSPHORYLATED (BY SIMILARITY).
CC -I SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
DR SWISS-2DPAGE: P99027; MOUSE.
KW Ribosomal protein; Phosphorylation.
FT MONTER 10 10
SQ SEQUENCE 10 AA; 1186 MW; 07121E3B45BDC2DB CRC64;

Query Match 25.0%; Score 2; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sy 6
DB 6 sy 7

RESULT 10
ID UR2A_CATCO STANDARD; PRT; 12 AA.
AC P04558;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE UROTENSIN IIA (U-IIA).
OS Catostomus commersoni (White sucker).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Catostomidae; Catostomus.
OX NCBI_TaxId=7971;
RN [1]
RP SEQUENCE.
RX MEDLINE=84041959; PubMed=6138758;
RA McMaster D., Lederis K.;
RT "Isolation and amino acid sequence of two urotensin II peptides from
RT Catostomus commersoni urophyses.";
RL Peptides 4:367-373(1983).
CC -I FUNCTION: UROTENSIN IS FOUND IN THE TELEOST CAUDAL NEUROSECRETORY
CC SYSTEM. IT HAS A SUGGESTED ROLE IN OSMOREGULATION AND AS A
CC CORTICOTROPIN-RELEASING FACTOR.
CC -I SIMILARITY: BELONGS TO THE UROTENSIN 2 FAMILY.
DR PIR: J50423; J50423.
DR InterPro: IPR001483; Urotensin_II.
DR Pfam: PF02083; Urotensin_II; 1.
DR PROSITE: PS00984; UROTENSIN_II; 1.
KW Hormone.
FT DISULFID
SQ SEQUENCE 12 AA; 1336 MW; 969C76DBB879CEBA CRC64;

Query Match 25.0%; Score 2; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wk 2
DB 8 wk 9

RESULT 11
ID UR2B_CATCO STANDARD; PRT; 12 AA.
AC P04559;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE UROTENSIN IIB (U-IIB).
OS Catostomus commersoni (White sucker).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;

```

```

OC Cypriniformes; Catostomidae; Catostomus.
OX NCBI_TaxId=7971;
RN [1]
RP SEQUENCE.
RX MEDLINE=84041959; PubMed=6138758;
RA McMaster D., Lederis K.;
RT "Isolation and amino acid sequence of two urotensin II peptides from
RT Catostomus commersoni urophyses.";
RL Peptides 4:367-373(1983).
CC -I FUNCTION: UROTENSIN IS FOUND IN THE TELEOST CAUDAL NEUROSECRETORY
CC SYSTEM. IT HAS A SUGGESTED ROLE IN OSMOREGULATION AND AS A
CC CORTICOTROPIN-RELEASING FACTOR.
CC -I SIMILARITY: BELONGS TO THE UROTENSIN 2 FAMILY.
DR PIR: J50424; J50424.
DR InterPro: IPR001483; Urotensin_II.
DR Pfam: PF02083; Urotensin_II; 1.
DR PROSITE: PS00984; UROTENSIN_II; 1.
KW Hormone.
FT DISULFID
SQ SEQUENCE 12 AA; 1437 MW; 73961BDBB879CEBB CRC64;

Query Match 25.0%; Score 2; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wk 2
DB 8 wk 9

RESULT 12
ID UR2B_CYPCA STANDARD; PRT; 12 AA.
AC P04561;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE UROTENSIN II-BETA.
OS Cyprinus carpio (Common carp).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Cyprininae; Cyprinus.
OX NCBI_TaxId=7962;
RN [1]
RP SEQUENCE.
RA Munekeata E., Ohtaki T., Ichikawa T., McMaster D., Lederis K.;
RL (in) Rich D.H., Gross E. (eds.);
RL Proceedings of the 7th american peptide symposium, pp. 69-72,
RL Pierce Chemical Co., Rockford IL. (1981).
CC -I FUNCTION: UROTENSIN IS FOUND IN THE TELEOST CAUDAL NEUROSECRETORY
CC SYSTEM. IT HAS A SUGGESTED ROLE IN OSMOREGULATION AND AS A
CC CORTICOTROPIN-RELEASING FACTOR.
CC -I SIMILARITY: BELONGS TO THE UROTENSIN 2 FAMILY.
DR InterPro: IPR001483; Urotensin_II.
DR Pfam: PF02083; Urotensin_II; 1.
DR PROSITE: PS00984; UROTENSIN_II; 1.
KW Hormone.
FT DISULFID
FT VARIANT 6 11 G->S.
SQ SEQUENCE 12 AA; 1407 MW; 73960A9FB879CEBB CRC64;

Query Match 25.0%; Score 2; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wk 2
DB 8 wk 9

RESULT 13

```

```

UR2_GILMI
ID UR2_GILMI STANDARD: PRT: 12 AA.
AC P01147;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE UROTENSIN II (U-II) (UII).
OS Gallichthys mirabilis (Long-jawed mudsnaker).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Acanthomorphi; Acanthopterygii; Percomorpha; Perciformes; Gobioidel;
OC Gobiidae; Gallichthys.
NCBI_TaxID=8222;
RX MEDLINE:81054904; PubMed=6107911;
RA Pearson D., Shively J.E., Clark B.R., Geschwind I.I., Barkley M.,
RT "Urotensin II: a somatostatin-like peptide in the caudal
neurosecretory system of fishes."
Proc. Natl. Acad. Sci. U.S.A. 77:5021-5024(1980).
-1- FUNCTION: UROTENSIN IS FOUND IN THE TELEOST CAUDAL NEUROSECRETORY
SYSTEM. IT HAS A SUGGESTED ROLE IN OSMOREGULATION AND AS A
CORTICOTROPIN-RELEASING FACTOR.
CC -1- SIMILARITY: BELONGS TO THE UROTENSIN 2 FAMILY.
DR InterPro: IPR001483; Urotensin_II.
DR Pfam: PF02083; Urotensin_II; 1.
DR PROSITE: PS00984; UROTENSIN_II; 1.
KW Hormone.
FT DISULFID 6 11
SQ SEQUENCE 12 AA: 1364 MW: 96868982679CEBA CRC64;

Query Match 25.0%; Score 2; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wk 2
DB 8 WK 9

RESULT 14
UR2_POLSP
ID UR2_POLSP STANDARD: PRT: 12 AA.
AC P81022;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE UROTENSIN II (U-II) (UII).
OS Polyodon spatula (North American paddlefish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Chondrostei; Acipenseriformes; Polyodontidae;
OC Polyodon.
NCBI_TaxID=7913;
RX MEDLINE:96051494; PubMed=8536944;
RA Waugh D., Youson J., Mims S.D., Sower S., Conlon J.M.;
RT "Urotensin II from the river lamprey (Lampetra fluviatilis), the sea
lamprey (Petromyzon marinus), and the paddlefish (Polyodon
spatula)."
Gen. Comp. Endocrinol. 99:323-332(1995).
-1- FUNCTION: HAS A SUGGESTED ROLE IN OSMOREGULATION AND AS A
CORTICOTROPIN-RELEASING FACTOR. PROBABLY INVOLVED IN SMOOTH
MUSCLE STIMULATION.
CC -1- SIMILARITY: BELONGS TO THE UROTENSIN 2 FAMILY.
DR InterPro: IPR001483; Urotensin_II.
DR Pfam: PF02083; Urotensin_II; 1.
DR PROSITE: PS00984; UROTENSIN_II; 1.
KW Hormone.

```

```

FT DISULFID 6 11 BY SIMILARITY
SQ SEQUENCE 12 AA: 1410 MW: 7551E9DB879CEBB CRC64;

Query Match 25.0%; Score 2; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wk 2
DB 8 WK 9

RESULT 15
UR2_SCYCA
ID UR2_SCYCA STANDARD: PRT: 12 AA.
AC P35490;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE UROTENSIN II (U-II) (UII).
OS Scyliorhinus canicula (Spotted dogfish) (Spotted catshark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphi; Galeoidea; Carcharhiniformes;
OC Scyliorhinidae; Scyliorhinus.
NCBI_TaxID=7830;
RX MEDLINE:92319231; PubMed=1620290;
RA Conlon J.M., O'Harte F., Smith D.D., Balment R.J., Hazon N.;
RT "Purification and characterization of urotensin II and parvalbumin
from an elasmobranch fish, Scyliorhinus canicula (common dogfish)."
Neuroendocrinology 55:230-235(1992).
CC -1- FUNCTION: HAS A SUGGESTED ROLE IN OSMOREGULATION AND AS A
CORTICOTROPIN-RELEASING FACTOR. PROBABLY INVOLVED IN SMOOTH
MUSCLE STIMULATION.
CC -1- SIMILARITY: BELONGS TO THE UROTENSIN 2 FAMILY.
DR InterPro: IPR001483; Urotensin_II.
DR Pfam: PF02083; Urotensin_II; 1.
DR PROSITE: PS00984; UROTENSIN_II; 1.
KW Hormone.
FT DISULFID 6 11
SQ SEQUENCE 12 AA: 1526 MW: 804729F9D579CEBA CRC64;

Query Match 25.0%; Score 2; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wk 2
DB 8 WK 9

RESULT 16
MLA_ANOCA
ID MLA_ANOCA STANDARD: PRT: 13 AA.
AC P41589;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MELANOTROPIN ALPHA (ALPHA-MSH).
OS Anolis carolinensis (Green anole) (American chameleon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauroi; Squamata; Iguania; Iguanidae; Polychrotinae; Anolis.
NCBI_TaxID=28377;
RX MEDLINE:92270473; PubMed=1667689;
RA Dorez R.M., Lancha A., Rand-Weaver M., Jankelew L., Adamczyk D.L.;
RT "Detection of a novel sequence change in the major form of alpha-MSH

```

RT isolated from the intermediate pituitary of the reptile, Anolis
 RT carolinensis.;
 RL Peptides 12:1261-1266(1991).
 CC -I- SIMILARITY: BELONGS TO THE POMC FAMILY.
 KW Hormone; Amidation.
 FT MOD_RES 13
 SO SEQUENCE 13 AA: 1608 MW: FP990A7358BB09C1 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 13;
 Best Local Similarity 100.0%; Pred. No. 1.3e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
 ||
 Db 1 sy 2

RESULT 17
 MLA_CAMDR STANDARD; PRT; 13 AA.
 AC P01198;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, last sequence update)
 DT 20-AUG-2001 (Rel. 40, last annotation update)
 DE MELANOTROPIN ALPHA (ALPHA-MSH). (Arabian camel), and
 OS Camelus dromedarius (Dromedary).
 CC Equus caballus (Horse).
 CC Eukaryota; Chordata; Craniala; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
 OX NCBI_TaxID=9838, 9796;
 RN [1]
 RP SEQUENCE.
 RC SPECIES=C.dromedarius;
 RX MEDLINE=7514634; PubMed=1125179;
 RA Li C.H., Danho W.O., Chung D., Rao A.J.;
 RT "Isolation, characterization, and amino acid sequence of
 melanotropins from camel pituitary glands.";
 RL Biochemistry 14:947-952(1975).
 RN [2]
 RP SEQUENCE.
 RC SPECIES=Horse; TISSUE=pituitary;
 RA Dixon J.S., Li C.H.;
 RT "The isolation and structure of alpha-melanocyte-stimulating hormone
 from horse pituitaries";
 RL J. Am. Chem. Soc. 82:4568-4572(1960).
 CC -I- SIMILARITY: BELONGS TO THE POMC FAMILY.
 DR PIR: A01464; MTCMAD.
 DR PIR: A91785; MTHOAD.
 KW Hormone; Acetylation; Amidation.
 FT MOD_RES 1
 FT MOD_RES 13
 FT MOD_RES 13
 SO SEQUENCE 13 AA: 1624 MW: FP991CA958BB09C1 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 13;
 Best Local Similarity 100.0%; Pred. No. 1.3e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
 ||
 Db 1 sy 2

RESULT 18
 MAST_PARID STANDARD; PRT; 14 AA.
 AC P42716;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, last sequence update)
 DT 01-NOV-1995 (Rel. 32, last annotation update)
 DE MASTOPARAN.

OS Parapolybia indica.
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Aculeata;
 OC Vespoidea; Vespidae; Polistinae; Parapolybia.
 OX NCBI_TaxID=31921;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Venom;
 RA Toki T., Yasuhara T., Nakajima T.;
 RT "Isolation and sequential analysis of peptides on the venom sac of
 Parapolybia indica.";
 RL Eisel Dobutsu 39:105-111(1989).
 CC -I- FUNCTION: MAST CELL DEGRANULATING PEPTIDE. ACTIVATES G PROTEINS
 CC THAT COUPLE TO PHOSPHOLIPASE C.
 KW Mast cell degranulation; Venom; Amidation.
 FT MOD_RES 14
 FT MOD_RES 14
 SO SEQUENCE 14 AA: 1619 MW: CA376CD3BA6D80DD CRC64;

Query Match 25.0%; Score 2; DB 1; Length 14;
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
 ||
 Db 3 WK 4

RESULT 19
 MAST_POLJA STANDARD; PRT; 14 AA.
 AC P01517;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, last sequence update)
 DT 01-AUG-1992 (Rel. 23, last annotation update)
 DE POLISTES MASTOPARAN.
 OS Polistes jadwigae (Paper wasp).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
 OC Vespoidea; Vespidae; Polistinae; Polistes.
 OX NCBI_TaxID=7457;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Venom;
 RA Hirai Y., Ueno Y., Yasuhara T., Yoshida H., Nakajima T.;
 RT "A new mast cell degranulating peptide, polistes mastoparan, in the
 venom of Polistes jadwigae.";
 RL Biomed. Res. 1:185-187(1980).
 CC -I- FUNCTION: MAST CELL DEGRANULATING PEPTIDE. ACTIVATES G PROTEINS
 CC THAT COUPLE TO PHOSPHOLIPASE C.
 DR PIR: A01780; OMNAPP.
 KW Mast cell degranulation; Venom; Amidation.
 FT MOD_RES 14
 FT MOD_RES 14
 SO SEQUENCE 14 AA: 1636 MW: 26472A53BF4778D8 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 14;
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
 ||
 Db 3 WK 4

RESULT 20
 MAST_VESXA STANDARD; PRT; 14 AA.
 AC P01515;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, last sequence update)
 DT 15-JUL-1999 (Rel. 38, last annotation update)
 DE MASTOPARAN X (MP-X).

```

OS Vespa xanthoptera (Japanese hornet).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
OC Vespoidea; Vespidae; Vespinae; Vespa.
OX NCBI_TaxID=7448;
RN [1]
RP SEQUENCE.
RC TISSUE=Venom;
RX MEDLINE=80155338; PubMed=540363;
RA Hirai Y., Kuwada M., Yasuhara T., Yoshida H., Nakajima T.;
RT "A new mast cell degranulating peptide homologous to mastoparan in
RL the venom of Japanese hornet (Vespa xanthoptera).";
RN Chem. Pharm. Bull. 27:1945-1946(1979).
RN [2]
RP STRUCTURE BY NMR.
RX MEDLINE=92304942; PubMed=1610813;
RA Wakamatsu K., Okada A., Miyazawa T., Ohya M., Higashijima T.;
RT "Membrane-bound conformation of mastoparan-X, a G-protein-activating
RL peptide.";
RN Biochemistry 31:5654-5660(1992).
RN [3]
RP STRUCTURE BY NMR.
RX MEDLINE=98206894; PubMed=9537994;
RA Kusunoki H., Wakamatsu K., Sato K., Miyazawa T., Kohno T.;
RT "G protein-bound conformation of mastoparan-X: heteronuclear
RN multidimensional transferred nuclear overhauser effect analysis of
RT peptide uniformly enriched with 13C and 15N.";
RN Biochemistry 37:4782-4790(1998).
CC -1- FUNCTION: MAST CELL DEGRANULATING PEPTIDE. ACTIVATES G PROTEINS
DR THAT COUPLE TO PHOSPHOLIPASE C.
DR PIR: A01778; OMVHX.
DR PDB: 1A13; 16-FEB-99.
DR Mast cell degranulation; Venom; Amidation; 3D-structure.
RN MOD_RES 14 14 AMIDATION.
FT MOD_RES 14 14 MW: 1557 MW; C85DE0D7AA7AB0DD CRC64;
SQ SEQUENCE 14 AA; 1557 MW; C85DE0D7AA7AB0DD CRC64;

Query Match 25.0%; Score 2; DB 1; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
DB 3 WK 4

RESULT 21
SMSL_MYOSC STANDARD; PRT; 14 AA.
ID SMSL_MYOSC
AC P20750;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SOMATOSTATIN I.
OS Myoxocephalus scorpius (Shorthorn sculpin) (Daddy sculpin),
OS Oncorhynchus kisutch (Coho salmon), and
OS Anguilla anguilla (European freshwater eel).
CC Actinopterygii; Neopterygii; Teleostei; Euteleostomi;
CC Acanthomorpha; Acanthopterygii; Percomorpha; Scorpaeniformes;
CC Cottidae; Cottidae; Myoxocephalus.
OX NCBI_TaxID=8097, 8019, 7936;
RN [1]
RP SEQUENCE.
RC SPECIES=M. scorpius; TISSUE=Pancreas;
RX MEDLINE=88029486; PubMed=2889597;
RA Conlon J.M., Davis M.S., Falkner S., Thim L.;
RT "Structural characterization of peptides derived from
RN prosomatostatin I and II isolated from the pancreatic islets of two
RT species of teleostean fish: the daddy sculpin and the flounder.";
RN Eur. J. Biochem. 168:647-652(1987).
RL [2]
RP SEQUENCE.

```

```

RC SPECIES=O. kisutch; TISSUE=Pancreas;
RX MEDLINE=87055212; PubMed=2877919;
RA Pilsetskaya E.M., Pollock H.G., Rouse J.B., Hamilton J.W.,
RA Kimmel J.R., Andrews P.C., Gorman A.;
RT "Characterization of coho salmon (Oncorhynchus kisutch) islet
RN somatostatin.";
RN Gen. Comp. Endocrinol. 63:252-263(1986).
RN [3]
RP SEQUENCE.
RC SPECIES=A. anguilla; TISSUE=Pancreas;
RX MEDLINE=89065329; PubMed=2904391;
RA Conlon J.M., Deacon C.F., Hazon N., Henderson I.W., Thim L.;
RT "Somatostatin-related and glucagon-related peptides with unusual
RN structural features from the European eel (Anguilla anguilla).";
RN Gen. Comp. Endocrinol. 72:181-189(1988).
CC -1- FUNCTION: SOMATOSTATIN INHIBITS THE RELEASE OF SOMATOTROPIN.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO THE SOMATOSTATIN FAMILY.
DR PIR: S00172; S00172.
DR PIR: B60842; B60842.
DR PIR: A60840; A60840.
KW Hormone; Multigene family.
FT DISULFID 3 14
SQ SEQUENCE 14 AA; 1640 MW; D6270F5C09682679 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
DB 8 WK 9

RESULT 22
SMS_ALMT STANDARD; PRT; 14 AA.
ID SMS_ALMT
AC P31885;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SOMATOSTATIN-14.
OS Alligator mississippiensis (American alligator), and
OS Trachemys scripta (Red-eared slider turtle) (Pseudemys scripta).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Crocodylidae; Alligatorinae; Alligator.
OX NCBI_TaxID=6496, 34903;
RN [1]
RP SEQUENCE.
RC SPECIES=A. mississippiensis; TISSUE=Stomach;
RX MEDLINE=93324451; PubMed=8101369;
RA Wang Y., Conlon J.M.;
RT "Neuroendocrine peptides (NPY, GRP, VIP, somatostatin) from the brain
RN and stomach of the alligator.";
RN Peptides 14:573-579(1993).
RN [2]
RP SEQUENCE.
RC SPECIES=T. scripta;
RX MEDLINE=90341082; PubMed=1974347;
RA Conlon J.M., Hicks J.W.;
RT "Isolation and structural characterization of insulin, glucagon and
RN somatostatin from the turtle, Pseudemys scripta.";
RN Peptides 11:461-466(1990).
CC -1- FUNCTION: SOMATOSTATIN INHIBITS THE RELEASE OF SOMATOTROPIN.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO THE SOMATOSTATIN FAMILY.
DR PIR: C60414; C60414.
KW Hormone.
FT DISULFID 3 14
SQ SEQUENCE 14 AA; 1640 MW; D6270F5C09682679 CRC64;

```

Query Match 25.0%; Score 2; DB 1; Length 14;
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wk 2
 11
 Db 8 wk 9

RESULT 23

CXA2_CONAL

ID CXA2_CONAL STANDARD; PRT; 15 AA.

AC P36640;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE ALPHA-CONOTOXIN AUIB.
 OS Conus aulicus (Court cone).
 OC Eukaryota; Metazoa; Mollusca; Gastropoda; Caenogastropoda;
 OC Neogastropoda; Conoidea; Conidae; Conus.
 OX NCBI_TaxID=89437;
 RN [1]
 RP SEQUENCE, AND SYNTHESIS.
 RC TISSUE-Venom;
 RX MEDLINE=99003392; PubMed=9786965;
 RA Luo S., Kulak J.M., Cartier G.E., Jacobsen R.B., Yoshikami D.,
 RA Olivera B.M., McIntosh J.M.;
 RT "Alpha-conotoxin AUIB selectively blocks alpha3 beta4 nicotinic
 acetylcholine receptors and nicotine-evoked norepinephrine release."
 RL J. Neurosci. 18:8571-8579(1998).
 CC -I- FUNCTION: ALPHA-CONOTOXINS ACT ON POSTSYNAPTIC MEMBRANES, THEY
 BIND TO THE NICOTINIC ACETYLCHOLINE RECEPTORS (NACHR) AND THUS
 INHIBIT THEM. THIS PEPTIDE BLOCKS MAMMALIAN NICOTINIC
 ACETYLCHOLINE RECEPTORS COMPOSED OF ALPHA-3/BETA-4 SUBUNITS.
 CC -I- MASS SPECTROMETRY: MW=1572.5; METHOD=ELECTROSPRAY.
 KW Postsynaptic neurotoxin; Acetylcholine receptor inhibitor; Amidation;
 KM Venom;
 FT DISULFID 2 8
 FT DISULFID 3 15
 FT MOD_RES 15 15
 FT MOD_RES 15 15
 SQ SEQUENCE 15 AA; 1578 MW; 84EFE95FDC700155 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 15;
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sy 6
 11
 Db 4 sy 5

RESULT 24

MALT_BACTQ

ID MALT_BACTQ STANDARD; PRT; 15 AA.

AC P80072;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-DEC-1992 (Rel. 24, Last annotation update)
 DE MALTASE (EC 3.2.1.20) (ALPHA-GLUCOSIDASE I) (FRAGMENT).
 OS Bacillus thermoamyliofaciens.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1425;
 RN [1]
 RP SEQUENCE.
 RC STRAIN=KP1071 / FERM P8477;
 RX MEDLINE=92209510; PubMed=1555585;
 RA Suzuki Y., Yonezawa K., Hattori M., Takai Y.;
 RT "Assignment of Bacillus thermoamyliofaciens KP1071
 alpha-glucosidase I to an exo-alpha-1,4-glucosidase, and its striking
 similarity to bacillary oligo-1,6-glucosidases in N-terminal sequence
 and in structural parameters calculated from the amino acid

RT composition."
 RL Eur. J. Biochem. 205:249-256(1992).
 CC -I- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING
 CC 1,4-LINKED D-GLUCOSE RESIDUES WITH RELEASE OF D-GLUCOSE.
 CC -I- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
 CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
 DR PIR: S21240; S21240.
 KW Hydrolase; Glycosidase.
 FT NON_TER 15 15
 SQ SEQUENCE 15 AA; 1929 MW; 62B4CE501F2D3042 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 15;
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wk 2
 11
 Db 6 wk 7

RESULT 25

PH2_PERAM

ID PH2_PERAM STANDARD; PRT; 15 AA.

AC P82695;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PEPTIDE HORMONE 2 (PEA-VEACID 2).
 OS Periplaneta americana (American cockroach).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
 CC Blattodea; Blattidae; Periplaneta.
 OX NCBI_TaxID=6978;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Abdominal perisymphathetic organs;
 RA Predel R.;
 RL Submitted (JUL-2000) to the SWISS-PROT data bank.
 CC -I- FUNCTION: UNKNOWN.
 KW Neuropeptide.
 SQ SEQUENCE 15 AA; 1603 MW; F353DC8B1F92B8BD CRC64;

Query Match 25.0%; Score 2; DB 1; Length 15;
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sy 6
 11
 Db 11 sy 12

Search completed: January 14, 2002, 08:08:21
 Job time: 706 sec

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Burns, Doane, Swecker & Mathis
;; STREET: P.O. Box 1404
;; CITY: Alexandria
;; STATE: Virginia
;; COUNTRY: United States
;; ZIP: 22313-1404
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/796,899
;; FILING DATE: 06-FEB-1997
;; CLASSIFICATION: 800
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/319,544
;; FILING DATE: 07-OCT-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Meuth, Donna M
;; REGISTRATION NUMBER: 36,607
;; REFERENCE/DOCKET NUMBER: 028754-005
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (703) 836-6620
;; TELEFAX: (703) 836-2021
;; INFORMATION FOR SEQ ID NO: 28:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 552 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; US-08-796-899-28

Query Match 37.5%; Score 3; DB 4; Length 552;
Best Local Similarity 100.0%; Pred. No. 6.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 8
Db 278 SFXG 281

RESULT 5
PCT-US95-03747-3
; Sequence 3, Application PC/TUS9503747
; GENERAL INFORMATION:
; APPLICANT: LA JOLLA CANCER RESEARCH FOUNDATION
; TITLE OF INVENTION: Brevican, A Glial Cell Proteoglycan
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/03747
; FILING DATE: 27-MAR-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Imbra, Richard J.
; REGISTRATION NUMBER: 37,643
; REFERENCE/DOCKET NUMBER: FP-LJ 1453
; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (619) 535-9001
;; TELEFAX: (619) 535-8949
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 908 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; PCT-US95-03747-3

Query Match 37.5%; Score 3; DB 5; Length 908;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 8
Db 171 SFXG 174

RESULT 6
US-08-889-841B-3
; Sequence 3, Application US/08889841B
; GENERAL INFORMATION:
; APPLICANT: Berman, Phillip W.
; TITLE OF INVENTION: HIV ENVELOPE POLYPEPTIDES AND VACCINE
; FILE REFERENCE: 14918-703CIP
; CURRENT APPLICATION NUMBER: US/08/889,841B
; CURRENT FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: US 60/676,737
; PRIOR FILING DATE: 1996-07-08
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 2
; TYPE: PRT
; ORGANISM: HIV
; US-08-889-841B-3

Query Match 25.0%; Score 2; DB 3; Length 2;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
Db 1 SF 2

RESULT 7
US-08-889-841B-6
; Sequence 6, Application US/08889841B
; GENERAL INFORMATION:
; APPLICANT: Berman, Phillip W.
; TITLE OF INVENTION: HIV ENVELOPE POLYPEPTIDES AND VACCINE
; FILE REFERENCE: 14918-703CIP
; CURRENT APPLICATION NUMBER: US/08/889,841B
; CURRENT FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: US 60/676,737
; PRIOR FILING DATE: 1996-07-08
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 2
; TYPE: PRT
; ORGANISM: HIV
; US-08-889-841B-6

Query Match 25.0%; Score 2; DB 3; Length 2;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6

```
Db      1 sf 2

RESULT 8
US-08-889-841B-13
; Sequence 13, Application US/08889841B
; GENERAL INFORMATION:
; APPLICANT: Berman, Phillip W.
; TITLE OF INVENTION: HIV ENVELOPE POLYPEPTIDES AND VACCINE
; FILE REFERENCE: 14918-703CIP
; CURRENT APPLICATION NUMBER: US/08/889,841B
; PRIOR FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: US 60/676,737
; PRIOR FILING DATE: 1996-07-08
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 2
; TYPE: PRT
; ORGANISM: HIV
US-08-889-841B-13

Query Match      25.0%; Score 2; DB 3; Length 2;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 sf 6
        ||
Db      1 sf 2

RESULT 9
US-08-889-841B-17
; Sequence 17, Application US/08889841B
; GENERAL INFORMATION:
; APPLICANT: Berman, Phillip W.
; TITLE OF INVENTION: HIV ENVELOPE POLYPEPTIDES AND VACCINE
; FILE REFERENCE: 14918-703CIP
; CURRENT APPLICATION NUMBER: US/08/889,841B
; PRIOR FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: US 60/676,737
; PRIOR FILING DATE: 1996-07-08
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 2
; TYPE: PRT
; ORGANISM: HIV
US-08-889-841B-17

Query Match      25.0%; Score 2; DB 3; Length 2;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 sf 6
        ||
Db      1 sf 2

RESULT 10
US-08-889-841B-20
; Sequence 20, Application US/08889841B
; GENERAL INFORMATION:
; APPLICANT: Berman, Phillip W.
; TITLE OF INVENTION: HIV ENVELOPE POLYPEPTIDES AND VACCINE
; FILE REFERENCE: 14918-703CIP
; CURRENT APPLICATION NUMBER: US/08/889,841B
; PRIOR FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: US 60/676,737
; PRIOR FILING DATE: 1996-07-08
```

```
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 20
; LENGTH: 2
; TYPE: PRT
; ORGANISM: HIV
US-08-889-841B-20

Query Match      25.0%; Score 2; DB 3; Length 2;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 sf 6
        ||
Db      1 sf 2

RESULT 11
US-08-889-841B-34
; Sequence 34, Application US/08889841B
; GENERAL INFORMATION:
; APPLICANT: Berman, Phillip W.
; TITLE OF INVENTION: HIV ENVELOPE POLYPEPTIDES AND VACCINE
; FILE REFERENCE: 14918-703CIP
; CURRENT APPLICATION NUMBER: US/08/889,841B
; PRIOR FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: US 60/676,737
; PRIOR FILING DATE: 1996-07-08
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 2
; TYPE: PRT
; ORGANISM: HIV
US-08-889-841B-34

Query Match      25.0%; Score 2; DB 3; Length 2;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 sf 6
        ||
Db      1 sf 2

RESULT 12
US-08-889-841B-37
; Sequence 37, Application US/08889841B
; GENERAL INFORMATION:
; APPLICANT: Berman, Phillip W.
; TITLE OF INVENTION: HIV ENVELOPE POLYPEPTIDES AND VACCINE
; FILE REFERENCE: 14918-703CIP
; CURRENT APPLICATION NUMBER: US/08/889,841B
; PRIOR FILING DATE: 1997-07-08
; PRIOR APPLICATION NUMBER: US 60/676,737
; PRIOR FILING DATE: 1996-07-08
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 37
; LENGTH: 2
; TYPE: PRT
; ORGANISM: HIV
US-08-889-841B-37

Query Match      25.0%; Score 2; DB 3; Length 2;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 sf 6
        ||
Db      1 sf 2
```


DB 1 SF 2

RESULT 13
US-08-889-841B-40
Sequence 40, Application US/08889841B
GENERAL INFORMATION:
APPLICANT: Berman, Phillip W.
TITLE OF INVENTION: HIV ENVELOPE POLYPEPTIDES AND VACCINE
FILE REFERENCE: 14918-703CIP
CURRENT APPLICATION NUMBER: US/08/889,841B
CURRENT FILING DATE: 1997-07-08
PRIOR APPLICATION NUMBER: US 60/676,737
NUMBER OF SEQ ID NOS: 57
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 40
LENGTH: 2
TYPE: PRT
ORGANISM: HIV
US-08-889-841B-40

Query Match 25.0%; Score 2; DB 3; Length 2;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sf 6
DB 1 SF 2

RESULT 14
US-09-461-697-405
Sequence 405, Application US/09461697
Patent No. 6277974
GENERAL INFORMATION:
APPLICANT: COGENT NEUROSCIENCE, Inc.
APPLICANT: Lo, Donald C.
APPLICANT: Barney, Shawn
APPLICANT: Thomas, Mary Beth
APPLICANT: Portbury, Stuart D.
APPLICANT: Puranam, Kasturi
APPLICANT: Katz, Lawrence C.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
FILE REFERENCE: 10001-005-999
CURRENT APPLICATION NUMBER: US/09/461,697
CURRENT FILING DATE: 1999-12-14
NUMBER OF SEQ ID NOS: 466
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 405
LENGTH: 3
TYPE: PRT
ORGANISM: Homo sapiens
US-09-461-697-405

Query Match 25.0%; Score 2; DB 4; Length 3;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wk 2
DB 2 WK 3

RESULT 15
US-07-657-769B-14
Sequence 14, Application US/07657769B
Patent No. 5256766
GENERAL INFORMATION:

APPLICANT: COUGHLIN, SHAWN R.
TITLE OF INVENTION: RECOMBINANT RECEPTOR AND RELATED
TITLE OF INVENTION: PHARMACEUTICALS
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: TRELL & MANELLA
STREET: 545 MIDDLEFIELD ROAD, SUITE 200
CITY: MENLO PARK
STATE: CA
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/657,769B
FILING DATE: 19910219
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 2000-0502.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-327-7250
TELEFAX: 415-327-2951
TELEX: 706141
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
US-07-657-769B-14

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sf 6
DB 1 SF 2

RESULT 16
US-07-714-540-11
Sequence 11, Application US/07714540
Patent No. 5262521
GENERAL INFORMATION:
APPLICANT: Almqvist, Ronald G.
APPLICANT: Toll, Lawrence
TITLE OF INVENTION: ISOLATED ATRIAL PEPTIDE-DEGRADING
TITLE OF INVENTION: ENZYME AND NOVEL COMPOUNDS USEFUL AS INHIBITORS THEREOF
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Irell & Manella
STREET: 545 Middlefield Road, Suite 200
CITY: Menlo Park
STATE: California
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/714,540
FILING DATE: 19910607
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:

```

? NAME: Reed, Diane E.
? REGISTRATION NUMBER: 31,292
? REFERENCE/DOCKET NUMBER: 8500-0135.00
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 415-327-7250
? TELEFAX: 415-327-2951
? TELERX: 706141
? INFORMATION FOR SEQ ID NO: 11:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 4 amino acids
? TYPE: AMINO ACID
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: protein
US-07-714-540-11

```

```

Query Match      25.0%; score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY	5	sf	6
Db	2	Sf	3

RESULT 17
US-07-714-540-12
; Sequence 12, Application US/07714540
; Patent No. 5262521
GENERAL INFORMATION:
APPLICANT: Almqvist, Ronald G.
APPLICANT: Toll, Lawrence
TITLE OF INVENTION: ISOLATED ATRIAL PEPTIDE-DEGRADING
TITLE OF INVENTION: ENZYME AND NOVEL COMPOUNDS USEFUL AS INHIBITORS THEREOF
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Itell & Manella
STREET: 545 Middlefield Road, Suite 200
CITY: Menlo Park
STATE: California
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/714,540
FILING DATE: 19910607
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Dianne E.
REGISTRATION NUMBER: 31,292
REFERENCE/DOCKET NUMBER: 8500-0135.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-327-7250
TELEFAX: 415-327-2951
TELEX: 706141
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-714-540-12.

Query^Match	25.0%;	Score 2;	DB 1;	Length 4;
Best Local Similarity	100.0%;	Pred. No. 1.6e+05;		
Matches + 2: Conservative	0;	Mismatches	0;	Indels

QY	5	SF	6
Db	1	SF	2

```

1      RESULT 18
2      US-07-828-450-38
3      : Sequence 38, Application US/07828450
4      : Patent No. 5434133
5      :
6      : GENERAL INFORMATION:
7      :
8      : APPLICANT: TANAKA, SHOJI
9      : APPLICANT: MINAMITAKE, YOSHIHARU
10     : APPLICANT: KITAJIMA, YASUO
11     : APPLICANT: FURUYA, MAYUMI
12     : APPLICANT: MATSUO, HISAYUKI
13     : TITLE OF INVENTION: CNP ANALOG PEPTIDES AND THEIR USE
14     :
15     : NUMBER OF SEQUENCES: 42
16     :
17     : CORRESPONDENCE ADDRESS:
18     : ADDRESSEE: CUSHMAN, DABRY & CUSHMAN
19     : STREET: 1625 L STREET, N.W.
20     : CITY: WASHINGTON
21     : STATE: D.C.
22     : COUNTRY: USA
23     : ZIP: 20036
24     :
25     : COMPUTER READABLE FORM:
26     : MEDIUM TYPE: Floppy disk
27     : COMPUTER: IBM PC compatible
28     : OPERATING SYSTEM: PC-DOS/MS-DOS
29     : SOFTWARE: Patent Release #1.0, Version #1.25
30     : CURRENT APPLICATION DATA:
31     : APPLICATION NUMBER: US/07/828,450
32     : FILING DATE: 19920131
33     :
34     : CLASSIFICATION: 530
35     :
36     : ATTORNEY/AGENT INFORMATION:
37     : NAME: SCOTT, WATSON T.
38     : REGISTRATION NUMBER: 26,581
39     : REFERENCE/DOCKET NUMBER: 9437/94133
40     : TELECOMMUNICATION INFORMATION:
41     : TELEPHONE: 202-861-3067
42     : TELEFAX: 202-822-0944
43     :
44     : TELEX: 6714627 CUSH
45     :
46     : INFORMATION FOR SEQ ID NO: 38:
47     : SEQUENCE CHARACTERISTICS:
48     : LENGTH: 4 amino acids
49     : TYPE: AMINO ACID
50     : STRANDEDNESS: single
51     : TOPOLOGY: linear
52     :
53     : MOLECULE TYPE: peptide
54     :
55     : US-07-828-450-38

```

```

Query Match      2;   Score 2;   DB 1;   Length 4;
Best Similarity 100.0%;   Pval No. 1.6e+05;
Best Local       2;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;
Matches          2;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

```

QY	5	SF	6
Db	2	SF	3

RESULT 19
 US-08-079-445-3
 ; Sequence 3, Application US/08079445
 ; Patent No. 5440016
 ; GENERAL INFORMATION:
 ; APPLICANT: Blondelle, Sylvie E.
 ; APPLICANT: Pinilla, Clemencia
 ; APPLICANT: Elchler, Jutta
 ; APPLICANT: Houghten, Richard A.
 ; TITLE OF INVENTION: PEPTIDES HAVING ANTI-MICROBIAL,
 ; TITLE OF INVENTION: HEMOLYTIC, ANTI-MELITTIN OR ANTI-ENZYMATIC ACTIVITY
 ; NUMBER OF SEQUENCES: 15

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: CAMPBELL AND FLORES
;; STREET: 4370 La Jolla Village Drive, Suite 700
;; CITY: San Diego
;; STATE: California
;; COUNTRY: United States of America
;; ZIP: 92122
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/079,445
;; FILING DATE: 18-JUN-1993
;; CLASSIFICATION: 514
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Campbell, Cathryn A.
;; REGISTRATION NUMBER: P-HP 9648
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (619) 535-9001
;; TELEFAX: (619) 535-8949
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 4 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;;
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 1
;; OTHER INFORMATION: /label= Xaa
;; OTHER INFORMATION: /note= "Xaa-Kfmcoc"
;;
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 4
;; OTHER INFORMATION: /label= Xaa
;; OTHER INFORMATION: /note= "Xaa-any amino acid"
;;
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 4
;; OTHER INFORMATION: /note= "C-terminal amino acid 1s
;; OTHER INFORMATION: /note= "amdated"
;;
;; US-08-079-445-3
;;
Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 wk 2
11
Db 2 wk 3
;;
RESULT 20
US-07-840-077A-7
Sequence 7, Application US/07840077A
Patent No. 5443816
GENERAL INFORMATION:
APPLICANT:
APPLICANT: Zamora, Paul O.
APPLICANT: Rhodes, Buck A.
TITLE OF INVENTION: Peptide-Metal Ion
TITLE OF INVENTION: Pharmaceutical Preparation and Method
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rhomed Incorporated
STREET: 4261 Balloon Park
CITY: Albuquerque
STATE: NM
COUNTRY: U.S.A.
ZIP: 87109-5802

;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb
;; MEDIUM TYPE: Storage
;; COMPUTER: IBM PC/XT/AT, IBM PS/2 or
;; COMPUTER: compatibles
;; OPERATING SYSTEM: PC-DOS or MS-DOS
;; SOFTWARE: WordPerfect 6.0a for Windows
;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/840,077A
;; FILING DATE: 20-FEB-1992
;; CLASSIFICATION:
;;
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 07/816,477
;; FILING DATE: 03-JAN-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Deborah A. Peacock
;; REGISTRATION NUMBER: 31,649
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (505) 242-9677
;; TELEFAX: (505) 243-2542
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 4 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: Peptide
;; HYPOTHETICAL: No
;; ANTI-SENSE: No
;;
;; US-07-840-077A-7
;;
Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 wk 2
11
Db 2 wk 3
;;
RESULT 21
US-08-127-904-11
Sequence 11, Application US/08127904
Patent No. 5470951
GENERAL INFORMATION:
APPLICANT: Eugene Roberts
TITLE OF INVENTION: Method For Antagonizing
TITLE OF INVENTION: Amnesic Effects of Amyloid n
TITLE OF INVENTION: Protein and Improving the
TITLE OF INVENTION: Quality of Life in individuals
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: City of Hope
STREET: 1500 East Duarte Road
CITY: Duarte
STATE: California
COUNTRY: United States of America
ZIP: 91010-0269
COMPUTER READABLE FORM:
MEDIUM TYPE: 3M Double Density 5 1/4" diskette
COMPUTER: Wang PC
OPERATING SYSTEM: MS DOS Version 3.20
SOFTWARE: Microsoft
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/127,904
FILING DATE: 29 September 1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA: No. 5470951e
ATTORNEY/AGENT INFORMATION:
NAME: Irons, Edward S.
REGISTRATION NUMBER: 16,541

REFERENCE/DOCKET NUMBER: No. 5470951e
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 783-6040
TELEFAX: (202) 783-6031
TELEX: No. 5470951e
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 4
TYPE: Amino Acid
STRANDEDNESS:
TOPOLOGY: Unknown
US-08-127-904-11

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sf 6
DB 1 sf 2

RESULT 22
US-08-127-904-12
Sequence 12, Application US/08127904
Patent No. 5470951
GENERAL INFORMATION:
APPLICANT: Eugene Roberts
TITLE OF INVENTION: Method For Antagonizing
TITLE OF INVENTION: Amnestic Effects of Amyloid n
TITLE OF INVENTION: Protein and Improving the
TITLE OF INVENTION: Quality of Life in Individuals
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: City of Hope
STREET: 1500 East Duarte Road
CITY: Duarte
STATE: California
COUNTRY: United States of America
ZIP: 91010-0269
COMPUTER READABLE FORM:
MEDIUM TYPE: 3M Double Density 5 1/4" diskette
COMPUTER: Wang PC
OPERATING SYSTEM: MS DOS Version 3.20
SOFTWARE: Microsoft
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/127,904
FILING DATE: 29 September 1993
CLASSIFICATION: A24
PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER: No. 5470951e
ATTORNEY/AGENT INFORMATION:
NAME: Irons, Edward S.
REGISTRATION NUMBER: 16,541
REFERENCE/DOCKET NUMBER: No. 5470951e
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 783-6040
TELEFAX: (202) 783-6031
TELEX: No. 5470951e
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 4
TYPE: Amino Acid
STRANDEDNESS:
TOPOLOGY: Unknown
US-08-127-904-12

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sf 6
DB 1 sf 2

RESULT 23
US-08-405-933-14
Sequence 14, Application US/08405933
Patent No. 5516889
GENERAL INFORMATION:
APPLICANT: Hollenberg, Morley D.
APPLICANT: Matsoukas, John M.
APPLICANT: Moore, Graham J.
TITLE OF INVENTION: Synthetic Thrombin Receptor Peptides
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: Washington & Prince Streets, P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: USA
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,933
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/080,643
FILING DATE: 21-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Dillahunty, Mary Ann
REGISTRATION NUMBER: 34,576
REFERENCE/DOCKET NUMBER: 028722-059
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)854-7400
TELEFAX: (415)854-8275
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-405-933-14

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sf 6
DB 1 sf 2

RESULT 24
US-08-461-611-12
Sequence 12, Application US/08461611
Patent No. 5635477
GENERAL INFORMATION:
APPLICANT: William Frank Degrado, Sharon Anne Jackson, Shaker Ahmed Mousa, Anju
APPLICANT: Parthasarathy, Michael Sworin, Maria Rafalski
TITLE OF INVENTION: Cyclic Compounds Useful as
TITLE OF INVENTION: Inhibitors of Platelet Glycoprotein IIb/IIIa
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: The DuPont Merck Pharmaceutical
ADDRESSEE: Company
STREET: 1007 Market Street, Legal Department

```

CITY: Wilmington
STATE: DE
COUNTRY: U.S.
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.50 Inch disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple Macintosh
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,611
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA: none
APPLICATION NUMBER: 08/038,448
FILING DATE: March 29, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Ferguson, Blair O
REGISTRATION NUMBER: 34,329
REFERENCE/DOCKET NUMBER: BP-6543-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-1676
TELEFAX: 302-892-8536
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
TOPOLOGY: circular
MOLECULE TYPE: peptide
ORIGINAL SOURCE: synthetic
FEATURE:
OTHER INFORMATION: Example Number 12a;
OTHER INFORMATION: GPIIb/IIIa Inhibitor
US-08-461-611-12

```

```

Query Match      25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      6 fxg 8
      |||
      1 fxg 3
Db

```

```

RESULT 25
US-07-789-184-109
Sequence 109, Application US/07789184
Patent No. 5688768
GENERAL INFORMATION:
APPLICANT: COUGHLIN, SHAUN R.
APPLICANT: SCARBOROUGH, ROBERT M.
TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND
NUMBER OF INVENTION: RELATED PHARMACEUTICALS
NUMBER OF SEQUENCES: 223
CORRESPONDENCE ADDRESS:
ADDRESS: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/789,184
FILING DATE: 19911107
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.

```

```

REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 22000-20502.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 34-0154
INFORMATION FOR SEQ ID NO: 109:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
US-07-789-184-109

```

```

Query Match      25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      5 sf 6
      ||
      1 sf 2
Db

```

Search completed: January 14, 2002, 07:57:31
Job time: 376 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:58:37 ; Search time 45.4 Seconds

(without alignments)
13.423 Million cell updates/sec

Title: 09-185908-1D

Perfect score: 8

Sequence: 1 wxxxxafgx 8

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 50 summaries

Database : PIR,68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37.5	31	2	B18955	class II histocomp
2	37.5	41	2	A60753	beta-lactamase (EC
3	37.5	122	2	A85797	hypothetical prote
4	37.5	260	2	F85630	hypothetical prote
5	37.5	691	2	E86025	hypothetical prote
6	37.5	879	2	G85739	hypothetical prote
7	25.0	3	2	S68328	blood cell protein
8	25.0	3	2	A61068	locustakinin - mlg
9	25.0	7	1	A61324	dermorphin - Rohde
10	25.0	7	2	S36662	dermorphin (Lys-7)
11	25.0	7	2	S21230	dermorphin (Tyr-4,
12	25.0	7	2	S33244	neuromodulatory pe
13	25.0	7	2	S33246	neuromodulatory pe
14	25.0	7	2	S68004	hucollin, 75k chain
15	25.0	8	2	S37141	rspa protein - Erw
16	25.0	8	2	A46306	spasmogenic toxin
17	25.0	8	2	JS0317	leucokinin VII - M
18	25.0	8	2	S66646	cardioacceleratory
19	25.0	9	2	S66607	quinoline 2-oxidor
20	25.0	9	2	S63491	dissimilatory sulf
21	25.0	9	2	S36898	ribosomal protein
22	25.0	9	2	A26363	cardioactive pepti
23	25.0	9	2	S39766	cardioactive pepti
24	25.0	9	2	S27233	cardioactive pepti
25	25.0	9	2	JN0027	[phe-6]-mosact - s
26	25.0	9	2	S39767	cardioactive pepti
27	25.0	10	1	RHLMGS	gonadotropin - se
28	25.0	10	2	S66458	ferredoxin - Rhizo
29	25.0	10	2	S70251	nitrogenase (EC 1.

30	2	25.0	10	2	C61440	polylacturonase
31	2	25.0	10	2	A49187	gonadotropin-relea
32	2	25.0	11	1	EEOCC	eledoisin - curled
33	2	25.0	11	1	EEOCC	eledoisin - musky
34	2	25.0	11	2	S07203	uperolesin - frog (
35	2	25.0	11	2	A35594	buccalin - Callfor
36	2	25.0	11	2	E57789	gallbladder stone
37	2	25.0	11	4	S52252	hypothetical prote
38	2	25.0	12	1	UOGM2	urotensin II - lon
39	2	25.0	12	2	S42765	urotensin II - tel
40	2	25.0	12	2	JS0423	urotensin II-A pep
41	2	25.0	12	2	S29479	hypothetical prote
42	2	25.0	12	2	S69123	proton-translocati
43	2	25.0	12	2	S01122	photosystem II 3.7
44	2	25.0	12	2	A35585	cytokinin-binding
45	2	25.0	12	2	JS0424	urotensin II-B pep
46	2	25.0	12	2	S47393	T-cell antigen rec
47	2	25.0	13	2	P00445	urotensin II - lau
48	2	25.0	13	2	S01119	photosystem II pro
49	2	25.0	13	2	PC1008	40k extracellular
50	2	25.0	13	2	S63492	dissimilatory sulf

ALIGNMENTS

RESULT 1
B18955 class II histocompatibility antigen HLA-DR beta chain - human (fragments)
C:Species: Homo sapiens (man)
C:Date: 05-Jun-1987 #sequence_revision 05-Jun-1987 #text_change 30-May-1997
C:Accession: B18955
R:Wiman, K.; Claesson, L.; Rask, L.; Tragarth, L.; Peterson, P.A.
Biochemistry 21, 5351-5358, 1982
A:Title: Purification and partial amino acid sequence of papain-solubilized class II
A:Reference number: A90463; MUID:83075335
A:Accession: B18955
A:Molecule type: protein
A:Residues: 1-31 <WIM>
C:Superfamily: class II histocompatibility antigen; immunoglobulin homology

Query Match 37.5% Score 3; DB 2; Length 31;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 5 afxg 8
| | | |
Db 17 AFxG 20

RESULT 2
A60753 beta-lactamase (EC 3.5.2.6) - lysobacter enzymogenes (fragment)
C:Species: lysobacter enzymogenes
C:Date: 14-May-1993 #sequence_revision 14-May-1993 #text_change 07-May-1999
C:Accession: A60753
R:Von Tigerstrom, R.G.; Boras, G.J.
J. Gen. Microbiol. 136, 521-527, 1990
A:Title: beta-lactamase of lysobacter enzymogenes: Induction, purification and charac
A:Reference number: A60753; MUID:90362037
A:Accession: A60753
A:Molecule type: protein
A:Residues: 1-41 <VON>
C:Keywords: antibiotic resistance; hydrolase; monomer

Query Match 37.5% Score 3; DB 2; Length 41;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 5 afxg 8
| | | |

Db 36 AFXG 39

RESULT 3
A:Accession: A85797
hypothetical protein yebf [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
A:Accession: A85797
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouzis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: A85797
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-122 <STO>
A:Cross-references: GB:AE005174; NID:g12515900; PIDN:AG56837.1; GSPDB:GN00145; UWGP:Z28
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: yebf

Query Match 37.5%; Score 3; DB 2; Length 122;
Best Local Similarity 100.0%; Pred. No. 5.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 afxg 8
||||
Db 10 AFXG 13

RESULT 4
A:Accession: F85630
hypothetical protein Z1373 [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
A:Accession: F85630
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouzis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: F85630
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-260 <STO>
A:Cross-references: GB:AE005174; NID:g12514217; PIDN:AG55506.1; GSPDB:GN00145; UWGP:Z13
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z1373

Query Match 37.5%; Score 3; DB 2; Length 260;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 afxg 8
||||
Db 4 AFXG 7

RESULT 5
A:Accession: E86025
hypothetical protein yhjG [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
A:Accession: E86025
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouzis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E86025

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-691 <STO>
A:Cross-references: GB:AE005174; NID:g12518239; PIDN:AG58665.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: yhjG
C:Superfamily: Escherichia coli hypothetical 75.1k protein (tref-kdkg region)

Query Match 37.5%; Score 3; DB 2; Length 691;
Best Local Similarity 100.0%; Pred. No. 2.8e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 afxg 8
||||
Db 268 AFXG 271

RESULT 6
A:Accession: G85739
hypothetical protein ydbH [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
A:Accession: G85739
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
Iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouzis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: G85739
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-879 <STO>
A:Cross-references: GB:AE005174; NID:g12515319; PIDN:AG56379.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: ydbH
C:Superfamily: Escherichia coli membrane protein ydbH

Query Match 37.5%; Score 3; DB 2; Length 879;
Best Local Similarity 100.0%; Pred. No. 3.4e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 afxg 8
||||
Db 739 AFXG 742

RESULT 7
A:Accession: S68328
blood cell protein A - Molgula manhattensis (fragment)
C:Species: Molgula manhattensis
C>Date: 19-Mar-1997 #sequence_revision 29-Aug-1997 #text_change 11-Jun-1999
A:Accession: S68328
R:Taylor, S.W.; Ross, M.M.; Waite, J.H.
Arch. Biochem. Biophys. 324, 228-240, 1995
A:Title: Novel 3,4-di- and 3,4,5-trihydroxyphenylalanine-containing polypeptides from
A:Reference number: S68325; MUID:96132650
A:Accession: S68328
A:Molecule type: protein
A:Residues: 1-3 <TAY>

Query Match 25.0%; Score 2; DB 2; Length 3;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 af 6
||

Db 2 Af 2

RESULT 8
A61068
Locustakinin - migratory locust
C:Species: Locusta migratoria (migratory locust)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 08-Dec-1995
A:Accession: A61068
R:Schooft, L.; Holman, G.M.; Proost, P.; Van Damme, J.; Hayes, T.K.; De Loof, A.
Regul. Pept. 37, 49-57, 1992
A:Title: Locustakinin, a novel myotropic peptide from Locusta migratoria, isolation, pri
A:Reference number: A61068; MUID:92262851

A:Accession: A61068
A:Molecule type: protein
A:Residues: 1-6 <SCH>
C:Keywords: amidated carboxyl end; cephalomyotropic peptide; neuropeptide
F:6/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 25.0%; Score 2; DB 2; Length 6;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
||
Db 1 Af 2

RESULT 9
A61324
dermorphin - Rohde's leaf frog
C:Species: Phyllomedusa rohdei (Rohde's leaf frog)
C:Date: 17-Jul-1994 #sequence_revision 05-Aug-1994 #text_change 07-May-1999
A:Accession: A61324
R:Montecucchi, P.C.; De Castiglione, R.; Erspamer, V.
Int. J. Pept. Protein Res. 17, 316-321, 1981
A:Title: Identification of dermorphin and Hyp(6)-dermorphin in skin extracts of the Braz
A:Reference number: A61324; MUID:82029915

A:Accession: A61324
A:Molecule type: protein
A:Residues: 1-7 <NON>
C:Superfamily: dermorphin precursor; dermorphin precursor amino-terminal homology
C:Keywords: amidated carboxyl end; D-amino acid; hydroxyproline; neuropeptide; skin
F:2/Modified site: D-alanine (Ala) #status experimental
F:6/Modified site: 4-hydroxyproline (Pro) (partial) #status experimental
F:7/Modified site: amidated carboxyl end (Ser) #status experimental

Query Match 25.0%; Score 2; DB 1; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
||
Db 2 Af 3

RESULT 10
S36662
dermorphin (lys-7) [validated] - two-colored leaf frog
C:Species: Phyllomedusa bicolor (two-colored leaf frog)
C:Date: 10-Dec-1993 #sequence_revision 13-Mar-1997 #text_change 18-Aug-2000
A:Accession: S36662
R:Milgogna, G.; Severini, C.; Simmaco, M.; Negri, L.; Falconieri Erspamer, G.; Kreil, G.
FEBS Lett. 302, 151-154, 1992
A:Title: Identification and characterization of two dermorphins from skin extracts of th
A:Reference number: S21152; MUID:92339502

A:Accession: S36662
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-7 <MG>
C:Superfamily: dermorphin precursor; dermorphin precursor amino-terminal homology

Query Match 25.0%; Score 2; DB 2; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
||
Db 2 Af 3

RESULT 11
S21230
dermorphin (Trp-4, Asn-7) [validated] - two-colored leaf frog (fragment)
C:Species: Phyllomedusa bicolor (two-colored leaf frog)
C:Date: 19-Mar-1997 #sequence_revision 10-Oct-1997 #text_change 18-Aug-2000
A:Accession: S21230
R:Milgogna, G.; Severini, C.; Simmaco, M.; Negri, L.; Falconieri Erspamer, G.; Kreil, G.
FEBS Lett. 302, 151-154, 1992
A:Title: Identification and characterization of two dermorphins from skin extracts of
A:Reference number: S21152; MUID:92339502

A:Accession: S21230
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-7 <MG>
C:Superfamily: dermorphin precursor; dermorphin precursor amino-terminal homology

Query Match 25.0%; Score 2; DB 2; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
||
Db 2 Af 3

RESULT 12
S33244
neuromodulatory peptide Wwamide-1 - giant African snail
C:Species: Achatina fulica (giant African snail)
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Jul-1997
A:Accession: S33244
R:Minakata, H.; Ikeda, T.; Muneoka, Y.; Kobayashi, M.; Nomoto, K.
FEBS Lett. 323, 104-108, 1993
A:Title: Wwamide-1, -2 and -3: novel neuromodulatory peptides isolated from ganglia o
A:Reference number: S33244; MUID:93265912

A:Accession: S33244
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-7 <MIN>

Query Match 25.0%; Score 2; DB 2; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
||
Db 1 wk 2

RESULT 13
S33246
neuromodulatory peptide Wwamide-3 - giant African snail
C:Species: Achatina fulica (giant African snail)
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Jul-1997
A:Accession: S33246
R:Minakata, H.; Ikeda, T.; Muneoka, Y.; Kobayashi, M.; Nomoto, K.
FEBS Lett. 323, 104-108, 1993
A:Title: Wwamide-1, -2 and -3: novel neuromodulatory peptides isolated from ganglia o
A:Reference number: S33244; MUID:93265912

A:Accession: S33246

A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-7 <MIN>

Query Match 25.0%; Score 2; DB 2; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wk 2
11
Db 1 wk 2

RESULT 14
S68004
hucolin, 75K chain - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 17-Mar-1999
C:Accession: S68004
R:Edgar, P.F.
FEBS Lett. 375, 159-161, 1995
A:Title: Hucolin, a new corticosteroid-binding protein from human plasma with structural
A:Reference number: S68004; MUID:96087107
A:Accession: S68004
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-7 <ED6>

Query Match 25.0%; Score 2; DB 2; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 af 6
11
Db 2 AF 3

RESULT 15
S37141
rpsa Protein - Erwinia chrysanthemi
C:Species: Erwinia chrysanthemi
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 08-Oct-1999
C:Accession: S37141
R:Douillie, A.; Toussaint, A.; Faelen, M.
submitted to the EMBL Data Library, August 1993
A:Description: Identification of the integration host factor genes of E. chrysanthemi.
A:Reference number: S37139
A:Accession: S37141
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-8 <DDU>
A:Cross-references: EMBL:X74750; NID:g399669; PIDN:CAAS2769.1; PID:g581108

Query Match 25.0%; Score 2; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 af 6
11
Db 2 AF 3

RESULT 16
A46306
spasmogenic toxin PNVI - spider (Phoneutria nigriventer) (fragment)
C:Species: Phoneutria nigriventer
C:Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 17-Mar-1999
C:Accession: A46306
R:Marangoni, S.; Borges, N.C.C.; Marangoni, R.A.; Antunes, E.; Vieira, C.A.; Novello, J.
Toxicol 31, 377-384, 1993

Query Match 25.0%; Score 2; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 af 6
11
Db 1 AF 2

RESULT 17
JS0317
leucokinin VII - Madeira cockroach
C:Species: Leucophaea maderae (Madeira cockroach)
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 20-Jun-2000
C:Accession: JS0317
R:Holman, G.M.; Cook, B.J.; Nachman, R.J.
Comp. Biochem. Physiol. C 88, 31-34, 1987
A:Title: Isolation, primary structure and synthesis of leucokinin VII and VIII: the
A:Reference number: JS0317
A:Accession: JS0317
A:Molecule type: protein
A:Residues: 1-8 <HOL>
C:Keywords: amidated carboxyl end; cephalomyotropic peptide
F:8/Modified site: amidated carboxyl end (Gly) #status experimental

A:Title: Biochemical characterization of a vascular smooth muscle contracting polypep
A:Reference number: A46306; MUID:93276438
A:Accession: A46306
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-8 <MAR>

Query Match 25.0%; Score 2; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 af 6
11
Db 2 AF 3

RESULT 18
S66646
cardioacceleratory protein 2b - tobacco hornworm
C:Species: Manduca sexta (tobacco hornworm)
C:Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 17-Mar-1999
C:Accession: S66646
R:Huesmann, G.R.; Cheung, C.C.; Lei, P.K.; Lee, T.D.; Swiderik, K.M.; Tublitz, N.J.
FEBS Lett. 371, 311-314, 1995
A:Title: Amino acid sequence of CAP(2b), an insect cardioacceleratory peptide from th
A:Reference number: S66646; MUID:96013159
A:Accession: S66646
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-8 <HUP>

Query Match 25.0%; Score 2; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 af 6
11
Db 3 AF 4

RESULT 19
S66607
guinoline 2-oxidoreductase beta chain - Comamonas testosteroni (fragment)
C:Species: Comamonas testosteroni

Query Match 25.0%; Score 2; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 af 6
11
Db 4 AF 5

RESULT 20
S66607
guinoline 2-oxidoreductase beta chain - Comamonas testosteroni (fragment)
C:Species: Comamonas testosteroni

Query Match 25.0%; Score 2; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 af 6
11
Db 4 AF 5

RESULT 21
S66607
guinoline 2-oxidoreductase beta chain - Comamonas testosteroni (fragment)
C:Species: Comamonas testosteroni

C>Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 17-Mar-1999
C:Accession: S66607
C:Species: *Schach, S.; Tshisuaka, B.; Fetzner, S.; Lligans, F.*
Eur. J. Biochem. 232, 536-544, 1995
A:Title: Quinolone 2-oxidoreductase and 2-oxo-1,2-dihydroquinoline 5,6-dioxygenase from
A:Reference number: S66606; MUID:96035889
A:Accession: S66607
A:Molecule type: protein
A:Residues: 1-9 <SCH>
A:Experimental source: strain 63

Query Match 25.0%; Score 2; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 af 6
||
Db 5 AF 6

RESULT 20
S63491
disulfamylatory sulfite reductase beta chain, membrane-bound - *Desulfovibrio desulfuricans*
C:Species: *Desulfovibrio desulfuricans*
C>Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 17-Mar-1999
C:Accession: S63491
R:Steuber, J.; Arendsen, A.F.; Hagen, W.R.; Kroneck, P.M.H.
Eur. J. Biochem. 233, 873-879, 1995

A:Title: Molecular properties of the disulfamylatory sulfite reductase from *Desulfovibrio*
A:Reference number: S63489; MUID:96085152
A:Accession: S63491
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-9 <STE>

Query Match 25.0%; Score 2; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 af 6
||
Db 1 AF 2

RESULT 21
S36898
ribosomal protein S8 - *Mycobacterium bovis* (fragment)
C:Species: *Mycobacterium bovis*
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 13-Jan-1995
C:Accession: S36898
R:Ohara, N.; Kimura, M.; Higashi, Y.; Yamada, T.
FEBS Lett. 331, 9-14, 1993

A:Title: Isolation and amino acid sequence of the 30S ribosomal protein S19 from *Mycobacterium bovis*
A:Reference number: S36887; MUID:94009653
A:Accession: S36898
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-9 <OHA>
C:Keywords: protein biosynthesis; ribosome

Query Match 25.0%; Score 2; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 af 6
||
Db 8 AF 9

RESULT 22

A26363
cardioactive peptide - green crab (fragment)
C:Species: *Carcinus maenas* (green crab, common shore crab)
C>Date: 25-Oct-1987 #sequence_revision 25-Oct-1987 #text_change 18-Jun-1993
C:Accession: A26363
R:Stangler, J.; Hildich, C.; Beyreuther, K.; Keller, R.
Proc. Natl. Acad. Sci. U.S.A. 84, 575-579, 1987
A:Title: Unusual cardioactive peptide (CCAP) from pericardial organs of the shore crab
A:Reference number: A26363
A:Accession: A26363
A:Molecule type: protein
A:Residues: 1-9 <STRA>

Query Match 25.0%; Score 2; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 af 6
||
Db 5 AF 6

RESULT 23
S39766
cardioactive peptide CCAP - yellow mealworm
C:Species: *Tenebrio molitor* (yellow mealworm)
C>Date: 07-Oct-1994 #sequence_revision 12-Apr-1996 #text_change 07-May-1999
C:Accession: S39766
R:Furuya, K.; Iiao, S.; Reynolds, S.E.; Ota, R.B.; Hackett, M.; Schooley, D.A.
Biol. Chem. Hoppe-Seyler 374, 1065-1074, 1993
A:Title: Isolation and identification of a cardioactive peptide from *Tenebrio molitor*
A:Reference number: S39766; MUID:94176032
A:Accession: S39766
A:Molecule type: protein
A:Residues: 1-9 <FUR>
C:Keywords: neuropeptide

Query Match 25.0%; Score 2; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 af 6
||
Db 5 AF 6

RESULT 24
S27233
cardioactive peptide CCAP - tobacco hornworm
C:Species: *Manduca sexta* (tobacco hornworm)
C>Date: 09-Jun-1994 #sequence_revision 12-Apr-1996 #text_change 12-Apr-1996
C:Accession: S27233
R:Chung, C.C.; Loi, P.K.; Sylwester, A.W.; Lee, T.D.; Tublitz, N.J.
FEBS Lett. 313, 165-168, 1992
A:Title: Primary structure of a cardioactive neuropeptide from the tobacco hawkmoth,
A:Reference number: S27233; MUID:93050243
A:Accession: S27233
A:Molecule type: protein
A:Residues: 1-9 <CHE>
C:Keywords: neuropeptide

Query Match 25.0%; Score 2; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 af 6
||
Db 5 AF 6

RESULT 25

JN0027

[Phe-6]-mosact - sea urchin (Clypeaster japonicus)

C:Species: Clypeaster japonicus

C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 20-Dec-1992

C:Accession: JN0027

R:Suzuki, N.; Kurita, M.; Yoshino, K.I.; Kajiwara, H.; Nomura, K.; Yamaguchi, M.

Zool. Sci. 4, 649-656, 1987

A:Title: Purification and structure of mosact and its derivatives from the egg jelly of

A:Reference number: JN0025

A:Accession: JN0027

A:Molecule type: protein

A:Residues: 1-9 <SUZ>

C:Comment: Mosact, one of several sperm activating peptides located in egg jelly, and it

Query Match 25.0%; Score 2; DB 2; Length 9;

Best Local Similarity 100.0%; Pred. No. 2.2e+05;

Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 af 6

DB 5 AF 6

Search completed: January 14, 2002, 07:58:37

Job time: 387 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 08:08:22 ; Search time 30.66 Seconds

(without alignments)
9.567 Million cell updates/sec

Title: 09-185908-1D

Perfect score: 8

Sequence: 1 wxxxxfxg 8

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size : 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

Database : SwissProt_39.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25.0	6	1	LOK1_LOCM1	P41491 locusta mig
2	25.0	7	1	ALL2_CARMA	P81805 carcinius ma
3	25.0	7	1	ALL3_CARMA	P81806 carcinius ma
4	25.0	7	1	ALL4_CARMA	P81807 carcinius ma
5	25.0	7	1	ALL5_CARMA	P81808 carcinius ma
6	25.0	7	1	WMA2_ACHFU	P35920 achalina fu
7	25.0	7	1	WMA3_ACHFU	P35921 achalina fu
8	25.0	8	1	ALL2_CARMA	P81815 carcinius ma
9	25.0	8	1	ALL7_CARMA	P81809 carcinius ma
10	25.0	8	1	ALL8_CARMA	P81811 carcinius ma
11	25.0	8	1	ALL9_CARMA	P81812 carcinius ma
12	25.0	8	1	LCK1_LEUMA	P21140 leucophaea
13	25.0	8	1	LCK7_LEUMA	P19889 leucophaea
14	25.0	8	1	NS3_MYCTU	P81152 mycobacteri
15	25.0	8	1	ORMY_ORCLI	P82455 orconectes
16	25.0	9	1	RS1_ERMCH	P37985 erwiniia chr
17	25.0	9	1	ALL0_CARMA	P81813 carcinius ma
18	25.0	9	1	ALL1_CARMA	P81814 carcinius ma
19	25.0	9	1	CCAF_CARMA	P38556 carcinius ma
20	25.0	9	1	MOSE_CUYJA	P19853 clypeaster
21	25.0	9	1	TKC1_CALVO	P41517 calliphora
22	25.0	10	1	GON1_PERMA	P04378 petromyzon
23	25.0	10	1	GON3_PERMA	P30948 petromyzon
24	25.0	10	1	O208_COMTE	P80465 comanonas t
25	25.0	10	1	UP11_CABEL	P55956 caenorhabd1
26	25.0	11	1	TKC2_CALVO	P41518 calliphora
27	25.0	11	1	TKN1_UPEIN	P83026 uperoleia i
28	25.0	11	1	TKN1_UPEIN	P08612 uperoleia r
29	25.0	11	1	TKN1_UPEIN	P01293 eladone mos
30	25.0	11	1	TKN1_UPEIN	P04558 calostomus
31	25.0	12	1	UR2A_CARCO	P04559 calostomus
32	25.0	12	1	UR2B_CARCO	P04561 cyprius ca
33	25.0	12	1	UR2_GILMI	P01147 gillichthys

ALIGNMENTS

RESULT 1	LOC1_LOCM1	STANDARD:	PRT:	6 AA.	P81022 polyodon sp
AC P41491:	LOC1_LOCM1				P35490 scylliorhinu
DT 01-NOV-1995 (Rel. 32, Created)					P80178 lymnaea sta
DT 01-NOV-1995 (Rel. 32, Last sequence update)					P80179 lymnaea sta
DT 15-JUL-1998 (Rel. 36, Last annotation update)					P80180 lymnaea sta
DE LOCUSTAKININ I.					P80181 lymnaea sta
OS Locusta migratoria (Migratory locust).					P80182 lymnaea sta
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;					P42716 parapolylbia
OC Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;					P01517 polistes ja
OC Acridomorpha; Acridoidea; Acrididae; Locusta.					P01515 vespa xanth
OX NCBI_Taxid=7004;					P20750 myoxocephal
RN [1]					P31985 alligator m
RP SEQUENCE.					P30039 homo sapien
RC TISSUE=Corpora cardiaca;					P55936 didelphis m
RX MEDLINE=92262851; PubMed=1585017;					P03057 escherichia
RA Schoofs L., Holman G.M., Proost P., van Damme J., Hayes T.K.,					P80072 bacillus th
RA de loof A.;					
RT "Locustakinin, a novel myotropic peptide from Locusta migratoria,					
RL Regul. Pept. 37:49-57(1992)."					
CC -I- FUNCTION: MYOTROPIC PEPTIDE. MAY BE IMPORTANT IN THE STIMULATION					
CC OF ION TRANSPORT AND INHIBITION OF DIURETIC ACTIVITY IN MALPIGHIAN					
CC TUBULES.					
CC PIR: A61068.					
DR Neuropeptide; Amidation.					
KW MOD_RRS					
FT MOD_RRS					
SQ SEQUENCE 6 AA; 654 MW; 686365A5B9CDB000 CRC64;					
Query Match	25.0%;	Score 2;	DB 1;	Length 6;	
Best Local Similarity	100.0%;	Pred. No. 1e+05;			
Matches 2;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY 5 af 6					
Db 1 Af 2					
RESULT 2	ALL2_CARMA	STANDARD:	PRT:	7 AA.	
ID ALL2_CARMA					
AC P81805;					
DT 30-MAY-2000 (Rel. 39, Created)					
DT 30-MAY-2000 (Rel. 39, Last sequence update)					
DT 30-MAY-2000 (Rel. 39, Last annotation update)					
DE CARCINUSMASTIN 2.					
OS Carcinus maenas (Common shore crab) (Green crab).					
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;					
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;					

```
OC Eubrachyura; Portunioidea; Portunidae; Carcinus.
OX NCBI_TaxID=6759;
RN [1]
RP SEQUENCE.
RC TISSUE=Cerebral ganglion, and Thoracic ganglion;
RX MEDLINE=98121193; PubMed=9461295;
RA Duvé H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RT "Isolation and identification of multiple neuropeptides of the
RT allatostatin superfamily in the shore crab Carcinus maenas.";
RL Eur. J. Biochem. 250:727-734(1997).
CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KM Neuropeptide; Amidation; Multigene family.
FT MOD_RES 7 AA; 770 MW; 672879CDBC85DDB70 CRC64;
SQ SEQUENCE 7 AA; 770 MW; 672879CDBC85DDB70 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 7;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
   ||
Db 4 AF 5

RESULT 3
ALL3_CARMA STANDARD; PRT; 7 AA.
ID ALL3_CARMA
AC P81806;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE CARCINUSTATIN 3.
OS Carcinus maenas (Common shore crab) (Green crab).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC Eubrachyura; Portunioidea; Portunidae; Carcinus.
OX NCBI_TaxID=6759;
RN [1]
RP SEQUENCE.
RC TISSUE=Cerebral ganglion, and Thoracic ganglion;
RX MEDLINE=98121193; PubMed=9461295;
RA Duvé H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RT "Isolation and identification of multiple neuropeptides of the
RT allatostatin superfamily in the shore crab Carcinus maenas.";
RL Eur. J. Biochem. 250:727-734(1997).
CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KM Neuropeptide; Multigene family.
FT MOD_RES 7 AA; 796 MW; 672879CDBC8476B70 CRC64;
SQ SEQUENCE 7 AA; 796 MW; 672879CDBC8476B70 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 7;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
   ||
Db 4 AF 5

RESULT 4
ALL4_CARMA STANDARD; PRT; 7 AA.
ID ALL4_CARMA
AC P81807;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CARCINUSTATIN 4.
OS Carcinus maenas (Common shore crab) (Green crab).
```

```
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC Eubrachyura; Portunioidea; Portunidae; Carcinus.
OX NCBI_TaxID=6759;
RN [1]
RP SEQUENCE.
RC TISSUE=Cerebral ganglion, and Thoracic ganglion;
RX MEDLINE=98121193; PubMed=9461295;
RA Duvé H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RT "Isolation and identification of multiple neuropeptides of the
RT allatostatin superfamily in the shore crab Carcinus maenas.";
RL Eur. J. Biochem. 250:727-734(1997).
CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KM Neuropeptide; Multigene family.
FT MOD_RES 7 AA; 782 MW; 672879CDBC8476A00 CRC64;
SQ SEQUENCE 7 AA; 782 MW; 672879CDBC8476A00 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 7;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
   ||
Db 4 AF 5

RESULT 5
ALL5_CARMA STANDARD; PRT; 7 AA.
ID ALL5_CARMA
AC P81808;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE CARCINUSTATIN 5.
OS Carcinus maenas (Common shore crab) (Green crab).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC Eubrachyura; Portunioidea; Portunidae; Carcinus.
OX NCBI_TaxID=6759;
RN [1]
RP SEQUENCE.
RC TISSUE=Cerebral ganglion, and Thoracic ganglion;
RX MEDLINE=98121193; PubMed=9461295;
RA Duvé H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RT "Isolation and identification of multiple neuropeptides of the
RT allatostatin superfamily in the shore crab Carcinus maenas.";
RL Eur. J. Biochem. 250:727-734(1997).
CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KM Neuropeptide; Amidation; Multigene family.
FT MOD_RES 7 AA; 781 MW; 672879CDBC8476420 CRC64;
SQ SEQUENCE 7 AA; 781 MW; 672879CDBC8476420 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 7;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
   ||
Db 4 AF 5

RESULT 6
WMA2_ACHFU STANDARD; PRT; 7 AA.
ID WMA2_ACHFU
AC P35920;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE CARCINUSTATIN 4.
```

```

DE WMAWIDE-2.
OS Achatina fulica (Giant African snail).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;
OC Achatinacea; Achatinidae; Achatina.
OX NCBI_TaxID=6530;
RN [1]
RP SEQUENCE.
RC TISSUE:Ganglion;
RA MEDLINE:93265912; Pubmed=8495720;
RA Minkata H., Ikeda T., Muneoka Y., Kobayashi M., Nomoto K.;
RT "Wamide-1, -2 and -3: novel neuromodulatory peptides isolated from
RL ganglia of the African giant snail, Achatina fulica.";
RL FEBS Lett. 323:104-108(1993).
KW Neuropeptide; Amidation.
FT MOD_RES 7
SQ SEQUENCE 7 AA; 964 MW; 7362D5B686D32310 CRC64;

Query Match      25.0%; Score 2; DB 1; Length 7;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
   11
Db 1 wk 2

RESULT 7
WMA3_ACHFV STANDARD; PRT; 7 AA.
AC P35921;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE WMAWIDE-3.
OS Achatina fulica (Giant African snail).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;
OC Achatinacea; Achatinidae; Achatina.
OX NCBI_TaxID=6530;
RN [1]
RP SEQUENCE.
RC TISSUE:Ganglion;
RA MEDLINE:93265912; Pubmed=8495720;
RA Minkata H., Ikeda T., Muneoka Y., Kobayashi M., Nomoto K.;
RT "Wamide-1, -2 and -3: novel neuromodulatory peptides isolated from
RL ganglia of the African giant snail, Achatina fulica.";
RL FEBS Lett. 323:104-108(1993).
DR PIR: S33244; S33244.
KW Neuropeptide; Amidation.
FT MOD_RES 7
SQ SEQUENCE 7 AA; 965 MW; 7362D5B69B132310 CRC64;

Query Match      25.0%; Score 2; DB 1; Length 7;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
   11
Db 1 wk 2

RESULT 8
AL12_CARMA STANDARD; PRT; 8 AA.
ID AL12_CARMA
AC P81815;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CARCINUSTATIN 12.
OS Carcinus maenas (Common shore crab) (Green crab).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;

```

```

OC Eubrachyura; Portunoidae; Portunidae; Carcinus.
OX NCBI_TaxID=6759;
RN [1]
RP SEQUENCE.
RC TISSUE:Cerebral ganglion, and Thoracic ganglion;
RA Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA Thorpe A.;
RT "Isolation and identification of multiple neuropeptides of the
RL allatostatin superfamily in the shore crab Carcinus maenas.";
RL Eur. J. Biochem. 250:727-734(1997).
CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
CC -1- FUNCTION: MAY ACT AS A NEUROTANSMITTER OR NEUROMODULATOR.
KW Neuropeptide; Multigene family.
FT MOD_RES 8
SQ SEQUENCE 8 AA; 913 MW; 672879CDB569AB7 CRC64;

Query Match      25.0%; Score 2; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
   11
Db 5 af 6

RESULT 9
ALL7_CARMA STANDARD; PRT; 8 AA.
ID ALL7_CARMA
AC P81809; P81810; P81804;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CARCINUSTATIN 7 [CONTAINS: CARCINUSTATIN 6; CARCINUSTATIN 1].
OS Carcinus maenas (Common shore crab) (Green crab).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC Eubrachyura; Portunoidae; Portunidae; Carcinus.
OX NCBI_TaxID=6759;
RN [1]
RP SEQUENCE.
RC TISSUE:Cerebral ganglion, and Thoracic ganglion;
RA MEDLINE:98121193; Pubmed=9461295;
RA Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA Thorpe A.;
RT "Isolation and identification of multiple neuropeptides of the
RL allatostatin superfamily in the shore crab Carcinus maenas.";
RL Eur. J. Biochem. 250:727-734(1997).
CC -1- FUNCTION: MAY ACT AS A NEUROTANSMITTER OR NEUROMODULATOR.
CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW Neuropeptide; Multigene family.
FT CHAIN 1
FT CHAIN 2
FT CHAIN 4
FT MOD_RES 8
SQ SEQUENCE 8 AA; 825 MW; 922879CDB4775BD CRC64;

Query Match      25.0%; Score 2; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
   11
Db 5 af 6

RESULT 10
ALL8_CARMA STANDARD; PRT; 8 AA.
ID ALL8_CARMA
AC P81811;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)

```

DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CARCINUSTATIN 8. (Common shore crab) (Green crab).
OS Carcinus maenas (Common shore crab) (Green crab).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
CC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC Eubrachyura; Portunoidae; Portunidae; Carcinus.
OX NCBI_TaxID=6759;
RN [1]
RP SEQUENCE.
RC TISSUE=Cerebral ganglion, and Thoracic ganglion;
RX MEDLINE=98121193; PubMed=9461295;
RA Duvé H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
Thorpé A.;
RT "Isolation and identification of multiple neuropeptides of the
allatostatin superfamily in the shore crab Carcinus maenas.";
RL Eur. J. Biochem. 250:727-734(1997).
CC -I- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC -I- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW Neuropeptide; Amidation; Multigene family.
FT MOD.RES 8
SQ SEQUENCE 8 AA; 795 MW; 922879CDBC847687D CRC64;

Query Match 25.0%; Score 2; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
II
DB 5 AF 6

RESULT 11
ID ALL9_CARMA STANDARD; PRT; 8 AA.
AC P81812;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CARCINUSTATIN 9.
OS Carcinus maenas (Common shore crab) (Green crab).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
CC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC Eubrachyura; Portunoidae; Portunidae; Carcinus.
OX NCBI_TaxID=6759;
RN [1]
RP SEQUENCE.
RC TISSUE=Cerebral ganglion, and Thoracic ganglion;
RX MEDLINE=98121193; PubMed=9461295;
RA Duvé H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
Thorpé A.;
RT "Isolation and identification of multiple neuropeptides of the
allatostatin superfamily in the shore crab Carcinus maenas.";
RL Eur. J. Biochem. 250:727-734(1997).
CC -I- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC -I- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW Neuropeptide; Amidation; Multigene family.
FT MOD.RES 8
SQ SEQUENCE 8 AA; 781 MW; 7C2879CDBC8476878 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
II
DB 5 AF 6
RESULT 12
ID LCK1_LEUMA STANDARD; PRT; 8 AA.

AC P21140;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-MAY-1991 (Rel. 18, Last annotation update)
DE LEUCOKININ I (L-I).
OS Leucophaea maderae (Madeira cockroach).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC Blaberoidea; Blaberidae; Leucophaea.
OX NCBI_TaxID=6988;
RN [1]
RP SEQUENCE, AND SYNTHESIS.
RC TISSUE=Head;
RA Holman G.M., Cook B.J., Nachman R.J.;
RT "Isolation, primary structure and synthesis of two neuropeptides
from Leucophaea maderae: members of a new family of
cephalomyotropic.";
RL Comp. Biochem. Physiol. 84C:205-211(1986).
CC -I- FUNCTION: THIS CEPHALOMYOTROPIC PEPTIDE STIMULATES CONTRACTILE
CC -I- ACTIVITY OF COCKROACH PROTODERM (HINDGUT).
CC -I- SIMILARITY: TO THE OTHER LEUCOKININS.
KW Neuropeptide; Amidation.
FT MOD.RES 8
SQ SEQUENCE 8 AA; 893 MW; DC6365B49DCD76A CRC64;

Query Match 25.0%; Score 2; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
II
DB 3 AF 4

RESULT 13
ID LCK7_LEUMA STANDARD; PRT; 8 AA.
AC P19989;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE LEUCOKININ VII (L-VII).
OS Leucophaea maderae (Madeira cockroach).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC Blaberoidea; Blaberidae; Leucophaea.
OX NCBI_TaxID=6988;
RN [1]
RP SEQUENCE.
RC TISSUE=Head;
RA Holman G.M., Cook B.J., Nachman R.J.;
RT "Isolation, primary structure and synthesis of leucokinin VII and
VIII: the final members of this new family of cephalomyotropic
peptides isolated from head extracts of Leucophaea maderae.";
RL Comp. Biochem. Physiol. 88C:31-34(1987).
CC -I- FUNCTION: THIS CEPHALOMYOTROPIC PEPTIDE STIMULATES CONTRACTILE
CC -I- ACTIVITY OF COCKROACH PROTODERM (HINDGUT).
CC -I- SIMILARITY: TO THE OTHER LEUCOKININS.
KW Neuropeptide; Amidation.
FT MOD.RES 8
SQ SEQUENCE 8 AA; 866 MW; DC6365B49DCD76A CRC64;

Query Match 25.0%; Score 2; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
II
DB 3 AF 4

RESULT 14
NS3_MYCTU STANDARD: PRT: 8 AA.
ID NS3_MYCTU
AC P81152;
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE 30 KDA NON-SECRETORY PROTEIN 3 (FRAGMENT).
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE.
RC STRAIN=H37RV;
RA Praead H.K., Annapurna P.S.;
RL Submitted (DEC-1997) to the SWISS-PROT data bank.
CC -1- CAUTION: WE ARE UNABLE TO FIND THIS PROTEIN IN THE TRANSLATION OF
FT THE GENOME OF STRAIN H37RV.
FT NON_TER 1 1
FT 8 8
SQ SEQUENCE 8 AA: 919 MW: 803DC40B19CDDC2D2 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
11
DB 3 af 4

RESULT 15
ORMY_ORCLI STANDARD: PRT: 8 AA.
ID ORMY_ORCLI
AC P82455;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ORCOMYOTROPIN (OMT).
OS Orcomyotropea limosus (Spirincheek crayfish).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC Astacoidae; Cambaridae; Orconectes.
OX NCBI_TaxID=28379;
RN [1]
RP SEQUENCE, MASS SPECTROMETRY, AND AMIDATION.
RC TISSUE=Hindgut;
RX MEDLINE=20411310; PubMed:10952880;
RA Dirksen H., Burdzik S., Sauter A., Keller R.;
RT "Two orcockins and the novel octapeptide orcomyotropin in the hindgut
of the crayfish Orcomyotropea limosus: Identified myostimulatory
neuropeptides originating together in neurones of the terminal
abdominal ganglion.";
J. Exp. Biol. 203:2807-2818(2000).
CC -1- FUNCTION: MYOTROPIC PEPTIDE, ENHANCES BOTH THE FREQUENCY AND
AMPLITUDE OF SPONTANEOUS HINDGUT CONTRACTIONS. IT IS SYNTHESIZED
BY ABDOMINAL GANGLIONIC NEURONS.
CC -1- MASS SPECTROMETRY: MW=904.8; METHOD: FAB.
KW Amidation; Neuropeptide.
FT MOD_RES 8
FT 8
SQ SEQUENCE 8 AA: 905 MW: 87C861B1A9CDDAA9 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
11
DB 3 af 4

RESULT 16
RSL_ERMCH STANDARD: PRT: 8 AA.
ID RSL_ERMCH
AC P37985;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE 30S RIBOSOMAL PROTEIN S1 (FRAGMENT).
GN RPSA.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=5556;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3937;
RA Douillie A., Toussaint A., Faelen M.;
RL Submitted (AUG-1993) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: BINDS MRNA; THUS FACILITATING RECOGNITION OF THE
CC INITIATION POINT. IT IS NEEDED TO TRANSLATE MRNA WITH A SHORT
CC SHINE-DALGARNO (SD) PURINE-RICH SEQUENCE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE S1P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X74750; CAA52769.1; .
DR PIR: S37141; S37141.
KM Ribosomal protein; Repeat; RNA-binding.
FT NON_TER 1 1
FT 1
SQ SEQUENCE 8 AA: 837 MW: 9E18733DC5B339CD CRC64;

Query Match 25.0%; Score 2; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
11
DB 1 af 2

RESULT 17
AL10_CARMA STANDARD: PRT: 9 AA.
ID AL10_CARMA
AC P81813;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CARCINOSTATIN 10.
OS Carcinus maenas (Common shore crab) (Green crab).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC Eubrachyura; Portunoidae; Portunidae; Carcinus.
OX NCBI_TaxID=6759;
RN [1]
RP SEQUENCE.
RC TISSUE=Cerebral ganglion, and Thoracic ganglion;
RX MEDLINE=98121193; PubMed=9461295;
RA Dve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA Thoirpe A.;
RT "Isolation and identification of multiple neuropeptides of the
RT allatostatin superfamily in the shore crab Carcinus maenas.";
RL Eur. J. Biochem. 250:727-734(1997).
CC -1- FUNCTION: MAY ACT AS A NEUROTANSMITTER OR NEUROMODULATOR.
CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.

KW Neuropeptide; Amidation; Multigene family.
FT MOD_RES 9 9 AMIDATION.
SQ SEQUENCE 9 AA: 963 MW: 372D79CDBA776C7 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 9;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
II
DB 6 AF 7

RESULT 18
ALL_CARMA STANDARD; PRT; 9 AA.
ID _ALL_CARMA
AC P81814;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CARCINUSTATIN 11.
OS Carcinus maenas (Common shore crab) (Green crab).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC Eubrachyura; Portunoidae; Portunidae; Carcinus.
OX NCBI_TaxID=6759;
RN [1]
RP SEQUENCE.
RC TISSUE=Cerebral ganglion, and Thoracic ganglion;
RX MEDLINE=98121193; PubMed=9461295;
RA Duvé H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaroš P.P.,
Thorpe A.;
RT "Isolation and identification of multiple neuropeptides of the
allatostatin superfamily in the shore crab Carcinus maenas.";
RL Eur. J. Biochem. 250:727-734(1997).
CC -I- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC -I- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW Neuropeptide; Amidation; Multigene family.
FT MOD_RES 9 9 AMIDATION.
SQ SEQUENCE 9 AA: 927 MW: 832D79CDBA46D861 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 9;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
II
DB 6 AF 7

RESULT 19
CCAP_CARMA STANDARD; PRT; 9 AA.
ID _CCAP_CARMA
AC P38556;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CARDIOACTIVE PEPTIDE (CCAP).
OS Carcinus maenas (Common shore crab) (Green crab).
OC Manduca sexta (Tobacco hawkmoth) (Tobacco Hornworm),
Tenebrio molitor (Yellow mealworm), and
OC Spodoptera eridania (Southern armyworm).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC Eubrachyura; Portunoidae; Portunidae; Carcinus.
OX NCBI_TaxID=6759, 7130, 7067, 37547;
RN [1]
RP SEQUENCE.
RC SPECIES=C. maenas; TISSUE=pericardial organs;
RA Stangier J., Hilbich C., Beyreuther K., Keller R.;
RT "Unusual cardioactive peptide (CCAP) from pericardial organs of the

RT shore crab Carcinus maenas.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:575-579(1987).

RN [2]
RP SEQUENCE.
RC SPECIES=M. sexta;
RX MEDLINE=93050243; PubMed=1426284;
RA Cheung C.C., Loi P.K., Sylvestre A.W., Lee T.D., Tshltz N.J.;
RT "Primary structure of a cardioactive neuropeptide from the tobacco
hawkmoth, Manduca sexta.";
RL FEBS Lett. 313:165-168(1992).

RN [3]
RP SEQUENCE.
RC SPECIES=T. molitor, and S. eridania; TISSUE=Head;
RX MEDLINE=94176032; PubMed=8129851;
RA Furuya K., Liao S., Reynolds S.E., Ota R.B., Hackett M.,
Schooley D.A.;
RT "Isolation and identification of a cardioactive peptide from Tenebrio
molitor and Spodoptera eridania.";
RL Biol. Chem. Hoppe-Seyler 374:1065-1074(1993).
CC -I- FUNCTION: THE EFFECT OF CCAP IS BOTH INC- AND CHRONOTROPIC.
CC -I- TISSUE SPECIFICITY: STORED IN PERICARDIAL ORGANS AND RELEASED
INTO THE HEMOLYMPH.
DR PIR: A26363; A26363.
KW PIR: S27233; S27233.
DR Neuropeptide; Amidation.
FT DISULFID 3 9
FT MOD_RES 9 9 AMIDATION.
SQ SEQUENCE 9 AA: 959 MW: C5A861A9CDD44EB9 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 9;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
II
DB 5 AF 6

RESULT 20
MOSF_CLYXA STANDARD; PRT; 9 AA.
ID _MOSF_CLYXA
AC P19853;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE [PHE-6]-MOSACT.
OS Clypeaster japonicus (Sand dollar).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidae; Euechinoidae; Gnathostomata; Clypeasteroidea;
OC Clypeasteridae; Clypeaster.
OX NCBI_TaxID=7644;
RN [1]
RP SEQUENCE.
RC TISSUE=Egg jelly;
RA Suzuki N., Kurita M., Yoshino K.I., Kajitara H., Nomura K.,
Yamaguchi M.;
RT "Purification and structure of mosact and its derivatives from the
egg jelly of the sea urchin Clypeaster japonicus.";
RL Zool. Sci. 4:649-656(1987).
CC -I- FUNCTION: STIMULATES SPERM RESPIRATION AND MOTILITY.
DR PIR: JN0027; JN0027.
SQ SEQUENCE 9 AA: 924 MW: 93245729CDC5BAB5 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 9;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
II
DB 5 AF 6

RESULT 21
 ID TKCI_CALVO STANDARD: PRT: 9 AA.
 AC P41517:
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE CALLITACHYKININ I.
 OS Calliphora vomitoria (Blue blowfly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Oestroidea; Calliphoridae; Calliphora.
 CC NCBI_TaxId=27454;
 RX MEDLINE=95075727; PubMed=7984492;
 RA Lundquist C.T., Clotens F.L., Holman G.M., Nichols R., Nachman R.J., Naessel D.R.,
 RA "Callitachykinin I and II, two novel myotropic peptides isolated from the blowfly, Calliphora vomitoria, that have resemblances to tachykinins."
 RT Peptides 15:761-768(1994).
 RL -1- FUNCTION: MYOACTIVE PEPTIDE.
 CC -1- SIMILARITY: SOME SIMILARITY TO TACHYKININS.
 KW Tachykinin; Neuropeptide; Amidation.
 FT MOD_RES 9
 SQ SEQUENCE 9 AA: 981 MW; 2417C8B59CD1B7 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 9;
 Best Local Similarity 100.0%; Pred. No. 1e+05;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
 11
 DB 4 af 5

RESULT 22
 ID GONI_PETMA STANDARD: PRT: 10 AA.
 AC P04378:
 DT 20-MAR-1987 (Rel. 04, Created)
 DT 20-MAR-1987 (Rel. 04, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE GONADOLIBERIN I (GONADOTROPIN-RELEASING HORMONE I) (GNRH-I) (LULIBERIN I).
 OS Petromyzon marinus (Sea lamprey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia; Petromyzontiformes; Petromyzontidae; Petromyzon.
 CC NCBI_TaxId=7757;
 RX MEDLINE=86168192; PubMed=3514603;
 RA Sherwood N.M., Sower S.A., Marshak D.R., Fraser B.A., Brownstein M.J.,
 RT "Primary structure of gonadotropin-releasing hormone from lamprey brain."
 RL J. Biol. Chem. 261:4812-4819(1986).
 CC -1- FUNCTION: STIMULATES THE SECRETION OF BOTH LUTEINIZING AND FOLLICLE-STIMULATING HORMONES.
 CC -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.
 DR PIR: A01412; RHLMS.
 DR InterPro: IPR002012; GNRH.
 DR Pfam: PF00446; GNRH; 1.
 DR PROSITE: PS00473; GNRH; 1.
 KW Hormone; Amidation; Hypothalamus.
 KW MOD_RES 1
 FT MOD_RES 1
 SQ SEQUENCE 10 AA: 1244 MW; 1E4B36237B1735AB CRC64;

Query Match 25.0%; Score 2; DB 1; Length 10;
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
 11
 DB 7 wk 8

RESULT 23
 ID GONI_PETMA STANDARD: PRT: 10 AA.
 AC P30948:
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE GONADOLIBERIN III (GONADOTROPIN-RELEASING HORMONE III) (GNRH-III) (LULIBERIN III).
 OS Petromyzon marinus (Sea lamprey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia; Petromyzontiformes; Petromyzontidae; Petromyzon.
 CC NCBI_TaxId=7757;
 RX MEDLINE=93178316; PubMed=8440174;
 RA Sower S.A., Chiang Y.-C., Lovas S., Conlon J.M.,
 RT "Primary structure and biological activity of a third gonadotropin-releasing hormone from lamprey brain."
 RL Endocrinology 132:1125-1131(1993).
 CC -1- FUNCTION: STIMULATES THE SECRETION OF BOTH LUTEINIZING AND FOLLICLE-STIMULATING HORMONES.
 CC -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.
 DR InterPro: IPR002012; GNRH.
 DR Pfam: PF00446; GNRH; 1.
 DR PROSITE: PS00473; GNRH; 1.
 KW Hormone; Amidation; Hypothalamus.
 KW MOD_RES 1
 FT MOD_RES 1
 SQ SEQUENCE 10 AA: 1277 MW; 284B36237A1F5A3 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 10;
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
 11
 DB 7 wk 8

RESULT 24
 ID Q2OB_COMTE STANDARD: PRT: 10 AA.
 AC P00465:
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE QUINOLINE 2-OXIDOREDUCTASE, BETA CHAIN (EC 1.3.99.17) (FRAGMENT).
 OS Comamonas testosteroni (Pseudomonas testosteroni).
 OC Bacteria; Proteobacteria; Beta subdivision; Comamonadaceae; Comamonas.
 CC NCBI_TaxId=285;
 RX MEDLINE=96035889; PubMed=7556204;
 RA Schach S., Tshisuaka B., Fetzner S., Lingens F.,
 RT Quinolone 2-oxidoreductase and 2-oxo-1,2-dihydroquinoline 5,6-dioxygenase from Comamonas testosteroni 63. The first two enzymes in quinoline and 3-methylquinoline degradation."
 RL Eur. J. Biochem. 232:536-544(1995).
 CC -1- FUNCTION: CONVERTS (3-METHYL-)-QUINOLINE TO (3-METHYL-)-2-OXO-1,2-DIHYDROQUINOLINE.

CC -1- CATALYTIC ACTIVITY: QUINOLINE + ACCEPTOR + H(2)O = ISOQUINOLIN-
CC 1(2H)-ONE + REDUCED ACCEPTOR.
CC -1- COFACTOR: FAD, MOLYBDENUM AND IRON-SULFUR.
CC -1- PATHWAY: FIRST STEP IN THE DEGRADATION OF QUINOLINE AND
CC (3-METHYL-)QUINOLINE.
CC -1- SUBUNIT: HETEROHEXAMER OF TWO ALPHA CHAINS, TWO BETA CHAINS, AND
CC TWO GAMMA CHAINS (PROBABLE).
CC Oxidoreductase: Flavoprotein; FAD; Molybdenum.
KM NON_TER 10 10
SQ SEQUENCE 10 AA: 1241 MW: C2E2C25D9DC769 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
||
Db 5 AF 6

RESULT 25
UP11_CAEEL STANDARD; PRT; 10 AA.
ID UP11_CAEEL
AC P55956;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE UNKNOWN PROTEIN FROM 2D-PAGE (SPOT 11) (FRAGMENT).
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE.
RC STRAIN=BRISTOL N2;
RX MEDLINE=97295299; PubMed=9150941;
RA Blin L., Heid H., Liberatori S., Geier G., Pallini V., Zwilling R.;
RT "Two-dimensional gel electrophoresis of Caenorhabditis elegans
homogenates and identification of protein spots by microsequencing.";
RL Electrophoresis 18:357-362(1997).
FT NON_TER 10 10
SO SEQUENCE 10 AA: 1102 MW: 15255F85B7287B14 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
||
Db 1 AF 2

Search completed: January 14, 2002, 08:08:22
Job time: 707 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 08:07:33 ; Search time 81.98 Seconds
(without alignments)
14.274 Million cell updates/sec

Title: 09-185908-1D
Perfect score: 8
Sequence: 1 wxxxxfxg 8

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 50 summaries

Database :

- 1: SP_ARCHAEA:*
- 2: SP_BACTERIA:*
- 3: SP_FUNGI:*
- 4: SP_HUMAN:*
- 5: SP_INVERTEBRATE:*
- 6: SP_MAMMAL:*
- 7: SP_MHC:*
- 8: SP_ORGANELLE:*
- 9: SP_PHAGE:*
- 10: SP_PLANT:*
- 11: SP RODENT:*
- 12: SP_VIRUS:*
- 13: SP_VERTEBRATE:*
- 14: SP_UNCLASSIFIED:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37.5	40	6	09TRU3	O9trm3 oryctolagus
2	37.5	65	5	09GXM2	O9gxm2 leishmania
3	37.5	66	5	09GXM1	O9gxm1 leishmania
4	37.5	82	8	09TC61	O9tc61 melanoplus
5	37.5	88	5	09NMG8	O9nmg8 leishmania
6	37.5	98	12	09QTM5	O9qtm5 le virus.
7	37.5	100	5	09G107	O9g107 leishmania
8	37.5	110	8	034865	O34865 lufjanus bu
9	37.5	130	8	09T7B4	O9t7b4 trinomys ih
10	37.5	131	8	09T220	O9t220 trinomys ih
11	37.5	149	2	09K4Q2	O9k4q2 unidentified
12	37.5	151	2	059551	O59551 mycoplasma
13	37.5	180	8	09G578	O9g578 xenotodon
14	37.5	185	2	09ZM44	O9zm44 wolbachia s
15	37.5	213	2	086976	O86976 desulfococc
16	37.5	305	8	09TE36	O9te36 dicrostrongy
17	37.5	311	8	095792	O95792 emola cyanu
18	37.5	311	8	09Z2G1	O9z2g1 phyliotlis o
19	37.5	326	8	09MRX6	O9mrx6 rhexia mari

20	3	37.5	346	8	003302	003302 campophilus
21	3	37.5	347	8	003340	003340 sphyradicus
22	3	37.5	347	8	003341	003341 sphyradicus
23	3	37.5	355	8	09TEF8	09tef8 oryctolagus
24	3	37.5	374	8	021210	021210 parachromis
25	3	37.5	379	8	021218	021218 parachromis
26	3	37.5	732	12	065089	065089 foot-and-mo
27	2	25.0	7	4	015903	015903 homo sapien
28	2	25.0	7	8	095945	095945 saccharomyc
29	2	25.0	8	2	009258	009258 synchococc
30	2	25.0	8	5	P82689	P82689 periplaneta
31	2	25.0	8	11	062721	062721 ratcus norv
32	2	25.0	8	11	09JLD7	09jld7 mesocricetu
33	2	25.0	8	12	098XR9	098xr9 human immun
34	2	25.0	9	2	045852	045852 clostridium
35	2	25.0	9	2	046179	046179 clostridium
36	2	25.0	9	4	014277	014277 homo sapien
37	2	25.0	9	4	09UQW0	09uqw0 homo sapien
38	2	25.0	9	4	09UMAO	09uma0 homo sapien
39	2	25.0	9	10	09AXH8	09axh8 mesembryant
40	2	25.0	9	12	085599	085599 moloney mur
41	2	25.0	9	12	090350	090350 hepatitis g
42	2	25.0	10	2	09R7J8	09r7j8 helicobacte
43	2	25.0	10	8	09XMB4	09xmb4 aegilops ta
44	2	25.0	10	11	09QVK8	09qvk8 mus sp. mep
45	2	25.0	11	2	047451	047451 escherichia
46	2	25.0	11	2	P95518	P95518 pasteurilla
47	2	25.0	11	4	060614	060614 homo sapien
48	2	25.0	11	7	077883	077883 oreochromis
49	2	25.0	11	7	077912	077912 oreochromis
50	2	25.0	11	7	09TOB3	09tob3 homo sapien

ALIGNMENTS

RESULT 1
ID O9TRU3 PRELIMINARY: PRT: 40 AA.
AC O9TRU3;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE FACTOR IX HEAVY CHAIN.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI-TaxID=9986;
RN [1]
RP SEQUENCE.
RX MEDLINE=92132840; PubMed=1776139;
RA Rao L.V., Hoang A.D.;
RT "Purification and characterization of rabbit factor IX and its
RT existence as a two-chain factor IX alpha in circulating plasma.";
RL Thromb. Res. 64:57-68(1991).
CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC HSSP: P00761; IMCT.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin. 1.
KW Hydrolyase; Serine protease.
SQ SEQUENCE 40 AA: 4159 MW: C8D8CCED3AC3CF92 CRC64;

Query Match 37.5%; Score 3; DB 6; Length 40;
Best Local Similarity 100.0%; Pred. No. 5.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 afxg 8
|||
Db 24 AFXG 27

RESULT 2

```
09GXM2
ID 09GXM2 PRELIMINARY; PRT; 65 AA.
AC 09GXM2;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE PROBABLE PRE-MRNA SPLICING FACTOR RNA HELICASE PRP16 (FRAGMENT).
GN LM12.336.
OS Leishmania major.
OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OC NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,
RA Oliver K.;
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL390114; CAB98464.2; -.
KW Helicase.
FT NON_TER 1 1
FT SEQUENCE 65 AA; 7105 MW; 865D6D7220703391 CRC64;

Query Match
Best Local Similarity 100.0%; Score 3; DB 5; Length 65;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 afxg 8
   ||||
   1 AFXG 4

RESULT 3
09GX01 PRELIMINARY; PRT; 66 AA.
AC 09GX01;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE PROBABLE REXANK (FRAGMENT).
GN LM12.295.
OS Leishmania major.
OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OC NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,
RA Oliver K.;
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL390114; CAC02145.1; -.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; ank; 2.
DR SMART; SM00248; ANK; 2.
DR PROSITE; PSS0088; ANK_REPEAT; 2.
DR PROSITE; PSS0297; ANK_REPEAT_REGION; 1.
FT NON_TER 1 1
FT NON_TER 66 66
FT SEQUENCE 66 AA; 6919 MW; 604760C6D9670F8C CRC64;

Query Match
Best Local Similarity 100.0%; Score 3; DB 5; Length 66;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 afxg 8
   ||||
   1 AFXG 17

RESULT 4
09TC61
```

```
ID 09TC61 PRELIMINARY; PRT; 82 AA.
AC 09TC61;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CYTOCHROME B (FRAGMENT).
GN CYTB.
OS Melanoplus flavids.
OC Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Orthopteroidea; Orthoptera; Cellifera;
OC Acridomorpha; Acridoidea; Acrididae; Melanoplus.
OC NCBI_TaxID=103641;
RN [1]
RP SEQUENCE FROM N.A.
RA Chapco W., Kuperus W.R., Litzenberger G.S.;
RT "Molecular phylogeny of melanopline grasshoppers (Orthoptera:
RT Acrididae). The genus Melanoplus.";
RL Ann. Entomol. Soc. Amer. 92:617-623(1999).
CC -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN (BY SIMILARITY).
CC -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME B/B6 FAMILY.
DR EMBL; AF145493; AAD56053.1; -.
DR InterPro; IPR000179; Cyt_b_b6.
DR Pfam; PF00033; cytochrome_b_N; 1.
KW Electron transport; Heme; Mitochondrion; Respiratory chain;
FT NON_TER 1 1
FT NON_TER 82 82
FT SEQUENCE 82 AA; 9244 MW; 4EDAD443120D5CA4 CRC64;

Query Match
Best Local Similarity 100.0%; Score 3; DB 8; Length 82;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 afxg 8
   ||||
   50 AFXG 53

RESULT 5
09NNG8 PRELIMINARY; PRT; 88 AA.
AC 09NNG8;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE PROBABLE HYPOTHETICAL 27.7 KDA PROTEIN (FRAGMENT).
GN LM15.205.
OS Leishmania major.
OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OC NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,
RA Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL160371; CAC00242.1; -.
FT NON_TER 1 1
FT NON_TER 88 88
FT SEQUENCE 88 AA; 9379 MW; 878CAE746830DA9D CRC64;

Query Match
Best Local Similarity 100.0%; Score 3; DB 5; Length 88;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

OY 5 afxg 8
||||
DB 11 AFXG 14

RESULT 6
Q9Q7W5 PRELIMINARY; PRT: 98 AA.
AC Q9Q7W5
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
ORF2 PROTEIN (FRAGMENT).
DE TT virus.
OC Viruses; ssDNA viruses; unclassified ssDNA viruses.
OX NCBI_TaxID=68887;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TTV WAM973;
RA Sumitani, Hijioka M., Matsushita M., Inokokusno G., WIdjaya A.,
Sumatraldi D., Kanal K., Ohta Y., Mishiro S.;
"TT virus (TTV) genotypes in native and non-native prostitutes of
Irian Jaya, Indonesia: implication for non-occupational
transmission." to the EMBL/GenBank/DBJ databases.
RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
EMBL: AB031728; BAA86191.1; -.
FT NON_TER 1 1
FT 98 98
SQ SEQUENCE 98 AA; 11089 MW; FD01AA6D7C29C04F CRC64;

Query Match 37.5%; Score 3; DB 12; Length 98;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 afxg 8
||||
DB 59 AFXG 62

RESULT 7
Q9GY07 PRELIMINARY; PRT: 100 AA.
AC Q9GY07
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PROBABLE HYPOTHETICAL 77.1 KD PROTEIN (FRAGMENT).
LM12.160.
GN Leishmania major.
OS Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,
Oliver K.;
Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL: AL390114; CAC02042.1; -.
DR InterPro: IPR002202; HMG-COA_red
DR PROSITE: PS00066; HMG_COA_REDUCTASE_1; UNKNOWN_1.
FT NON_TER 1 1
FT 100 100
SQ SEQUENCE 100 AA; 11265 MW; 47230C5CD5456206 CRC64;

Query Match 37.5%; Score 3; DB 5; Length 100;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WXXXA 5
|||||

DB 1 WXXXA 5

RESULT 8
Q34865 PRELIMINARY; PRT: 110 AA.
ID Q34865
AC Q34865;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CYTOCHROME B (FRAGMENT).
GN Cytb.
OS Lutjanus buccanella.
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AC Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorphi; Acanthopterygii; Percormorpha; Perciformes; Percoidae;
OC Lutjanidae; Lutjaninae; Lutjanus.
OX NCBI_TaxID=40504;
RN [1]
RP SEQUENCE FROM N.A.
RA Sarver S.K., Freshwater D., Walsh P.J.;
Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
RL -1- FUNCTION: COMPONENT OF THE UBIDUTOL-CYTOCHROME C REDUCTASE
COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: OX(2) + 2 FERRICYTOCHROME C = O + 2
FERROCYTOCHROME C.
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
BOUND TO THE PROTEIN (BY SIMILARITY).
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B/B6 FAMILY.
DR EMBL: U26959; AAR70175.1; -.
DR InterPro: IPR000179; Cyt_b_b6.
DR Pfam: PF00033; cytochrome_b_n; 1.
DR PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
KW Electron transport; Heme; Mitochondrion; Respiratory chain;
KW Transmembrane.
FT NON_TER 1 1
FT 110 110
SQ SEQUENCE 110 AA; 12530 MW; 29AA811BC7EE6458 CRC64;

Query Match 37.5%; Score 3; DB 8; Length 110;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 afxg 8
||||
DB 99 AFXG 102

RESULT 9
Q9T7B4 PRELIMINARY; PRT: 130 AA.
ID Q9T7B4
AC Q9T7B4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CYTOCHROME B (FRAGMENT).
GN Cytb OR Cyt B.
OS Trinomys lheringi.
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystriocognathi; Echimyidae; Trinomys.
OX NCBI_TaxID=42830;
RN [1]
RP SEQUENCE FROM N.A.
RA Lara M.C., Patton J.L.;
"Evolutionary diversification of spiny rats (genus Trinomys, Rodentia:
Echimyidae) in the Atlantic Forest of Brazil.";

```
RL Zool. J. Linn. Soc. 0:0-0(2000).
CC -I- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -I- COPACITOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN (BY SIMILARITY).
CC -I- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RISKE PROTEIN (BY SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE CYTOCHROME B/B6 FAMILY.
CC EMBL: AF194318; AAF16120.1; -.
DR InterPro: IPR000179; Cyt_b_b6.
DR Pfam: PF00033; cytochrome_b_N; 1.
KW Electron transport; Heme; Mitochondrion; Respiratory chain;
KW Transmembrane.
FT NON_TER 130
SQ SEQUENCE 130 AA; 14687 MW; 374EFC940B9EABE9 CRC64;

Query Match 37.5%; Score 3; DB 8; Length 130;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 afxg 8
| | | |
DB 127 AFXG 130

RESULT 10
O9T220 PRELIMINARY; PRT; 131 AA.
ID O9T220:
DC 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CYTOCHROME B (FRAGMENT).
GN CYTB OR CYT B.
OS Trinomys lheringi.
OS Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystriocognath; Echimyidae; Trinomys.
OX NCBI_TaxId=42830;
[1]
RN SEQUENCE FROM N.A.
RA Lara M.C., Patton J.L.;
RT "Evolutionary diversification of spiny rats (genus Trinomys, Rodentia:
RT Echimyidae) in the Atlantic Forest of Brazil.";
RL Zool. J. Linn. Soc. 0:0-0(2000).
CC -I- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -I- COPACITOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN (BY SIMILARITY).
CC -I- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RISKE PROTEIN (BY SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE CYTOCHROME B/B6 FAMILY.
CC EMBL: AF194320; AAF16122.1; -.
DR EMBL: AF194319; AAF16121.1; -.
DR InterPro: IPR000179; Cyt_b_b6.
DR Pfam: PF00033; cytochrome_b_N; 1.
KW Electron transport; Heme; Mitochondrion; Respiratory chain;
KW Transmembrane.
FT NON_TER 131
SQ SEQUENCE 131 AA; 14850 MW; F537AEFC940B9EAB CRC64;

Query Match 37.5%; Score 3; DB 8; Length 131;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 afxg 8
| | | |
DB 127 AFXG 130
```

```
DB 127 AFXG 130

RESULT 11
O9K402 PRELIMINARY; PRT; 149 AA.
ID O9K402:
AC O9K402;
DC 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE AMOA PROTEIN (FRAGMENT).
GN AMOA.
OS unidentified proteobacterium.
OC Bacteria; Proteobacteria; environmental samples.
OX NCBI_TaxId=2722;
[1]
RN SEQUENCE FROM N.A.
RA Speksnijder A.G.C.L., Kowalchuk G.A., Jong de S., Laanbroek H.J.;
RT "Recovery of a Nitrosomonas-like amoA sequence group from freshwater
RT habitats.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ388566; CAB96453.1; -.
DR InterPro: IPR003393; AMO.
DR Pfam: PF02461; AMO; 1.
FT NON_TER 1
FT NON_TER 149
SQ SEQUENCE 149 AA; 16673 MW; A35D692633C6BBE6 CRC64;

Query Match 37.5%; Score 3; DB 2; Length 149;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 afxg 8
| | | |
DB 104 AFXG 107

RESULT 12
O59551 PRELIMINARY; PRT; 151 AA.
ID O59551:
AC O59551;
DC 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ADENOSINETRIPHOSPHATASE (EC 3.6.1.3) (ADENILPYROPHOSPHATASE) (ATP
DE MONOPHOSPHATASE) (ATPASE) (FRAGMENT).
GN ATPA.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxId=2104;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=M129;
RC MEDLINE=93054343; PubMed=1429453;
RA Wenzel R., Pirkle E., Hermann R.;
RT "Construction of an EcoRI restriction map of Mycoplasma pneumoniae and
RT localization of selected genes.";
RL J. Bacteriol. 174:7289-7296(1992).
CC -I- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + PHOSPHATE.
CC -I- SIMILARITY: TO ATPASE ALPHA/BETA CHAINS FAMILY.
CC EMBL: X67653; CAA47895.1; -.
DR InterPro: IPR000194; ATPase_alpha_beta.
DR Pfam: PF00006; ATP-synt_ad; 1.
KW Hydrogen ion transport; Hydrolase.
FT NON_TER 1
FT NON_TER 151
SQ SEQUENCE 151 AA; 16671 MW; A7FD199618264466 CRC64;

Query Match 37.5%; Score 3; DB 2; Length 151;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 afxg 8
| | | |
DB 104 AFXG 107
```


Matches, 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 afxg 8
|||||
DB 84 AFXG 87

RESULT 13

09G578 ID PRELIMINARY; PRT; 180 AA.

AC 09G578;

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)

DE CYTOCHROME B (FRAGMENT).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorphi; Acanthopterygii; Perciformi; Atherinomorpha;

OC Belontiiformes; Belontiidae; Xenentodon.

OX NCBI_TaxID=129108;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=N26;

RX MEDLINE=20458311; PubMed=11005301;

RA Lovejoy N.R.;

RT "Reinterpreting recapitulation: systematics of needlefishes and their

allies.";

RL Evolution 54:1349-1362(2000).

CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE

COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A

RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL

COUPLED TO ATP SYNTHESIS (BY SIMILARITY).

CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY

BOUND TO THE PROTEIN (BY SIMILARITY).

CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B/B6 FAMILY.

CC EMBL: AF243891; AAG42985.1; -

CC InterPro: IPR000179; Cyt_b_b6.

CC Pfam: PF000033; cytochrome_b_n: 1.

CC PROSITE: PS00192; CYTOCHROME_B_HEME.1.

DR Electon transport; Heme; Mitochondrion; Respiratory chain;

KW Transmembrane.

FT NON_TER 1 180

FT SEQUENCE 180 AA; 20032 MW; 01B64B87C00C5BD6 CRC64;

SO

Query Match 37.5%; Score 3; DB 8; Length 180;

Best Local Similarity 100.0%; Pred. No. 2.2e+03;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 afxg 8
|||||

DB 86 AFXG 89

RESULT 14

09ZHA4 ID PRELIMINARY; PRT; 185 AA.

AC 09ZHA4;

DT 01-MAY-1999 (TREMBLrel. 10, Created)

DT 01-MAY-1999 (TREMBLrel. 10, last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)

DE OTHER SURFACE PROTEIN PRECURSOR (FRAGMENT).

WSP.

OC Wolbachia sp. wKAYLC.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Wolbachiae; Wolbachia.

OC NCBI_TaxID=77833;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=WKAYLC;

RA van Meer M.M.M., Witteveldt J., Stoutamer R.;

RT "Phylogeny of the arthropod endosymbiont Wolbachia based on the wsp

gene.";

RL Insect Mol. Biol. 0:0-0(1999).

DR EMBL: AF071927; AAC77407.1;

DR InterPro: IPR002566; Surface_Ag_msp4.

DR Pfam: PF01617; Surface_Ag_2; 1.

FT NON_TER 1 185

FT SEQUENCE 185 AA; 19990 MW; 1F1D66A2B72FAE62 CRC64;

SO

Query Match 37.5%; Score 3; DB 2; Length 185;

Best Local Similarity 100.0%; Pred. No. 2.2e+03;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 afxg 8
|||||

DB 83 AFXG 86

RESULT 15

086976 ID PRELIMINARY; PRT; 213 AA.

AC 086976;

DT 01-NOV-1998 (TREMBLrel. 08, Created)

DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)

DE DISMILITORY SULFITE REDUCTASE ALPHA SUBUNIT (FRAGMENT).

OS Desulfococcus multivorans.

OC Bacteria; Proteobacteria; delta subdivision; Desulfococcus.

OX NCBI_TaxID=897;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 33890;

RX MEDLINE=98269016; PubMed=9603890;

RA Wagner M., Roger A.J., Flax J.L., Brusseau G.A., Stahl D.A.;

RT "Phylogeny of dissimilatory sulfite reductases supports an early

origin of sulfate respiration.";

RL J. Bacteriol. 180:2975-2982(1998).

DR EMBL: U58126; AAC24101.1;

DR InterPro: IPR000660; Nlr_Sir.

DR Pfam: PF01077; Nlr_Sir.1.

DR PRINTS: PR00397; SIROHAEM.

FT NON_TER 1 213

FT SEQUENCE 213 AA; 23637 MW; 49EF93808BF46ACA CRC64;

SO

Query Match 37.5%; Score 3; DB 2; Length 213;

Best Local Similarity 100.0%; Pred. No. 2.6e+03;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 afxg 8
|||||

DB 22 AFXG 25

RESULT 16

09TEJ6 ID PRELIMINARY; PRT; 305 AA.

AC 09TEJ6;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)

DE CYTOCHROME B (FRAGMENT).

GN CYTB.

OS Dicrostonyx torquatus (Arctic lemming).

OC Mitochondrion.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Arvicolinae;
OC Dicrostonyx.
RN NCB1_TaxID=85952;
RX [1]
RP SEQUENCE FROM N.A.
RA Fedorov V.B., Goropashnaya A.V.;
RT "The importance of ice ages in diversification of Arctic collared
lemmings (Dicrostonyx): evidence from the mitochondrial cytochrome b
region."
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -I- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
BOUND TO THE PROTEIN (BY SIMILARITY).
CC -I- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CYTOCHROME C1 AND THE RISKE PROTEIN (BY SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE CYTOCHROME B/B6 FAMILY.
EMBL; AJ238425; CAB53232.1; -.
DR InterPro: IPR000179; Cyt_b_b6.
DR Pfam: PF00032; cytochrome_b_c1; 1.
DR Pfam: PF00033; cytochrome_b_n; 1.
DR PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
DR PROSITE: PS00193; CYTOCHROME_B_OO; UNKNOWN.1.
KW Electron transport; Heme; Mitochondrion; Respiratory chain;
FT TRANSMEMBRANE.
FT NON_TER 1
FT NON_TER 305
SQ SEQUENCE 305 AA; 34456 MW; 1F4CA31BA50A7239 CRC64;

Query Match 37.5%; Score 3; DB 8; Length 305;
Best Local Similarity 100.0%; Pred. No. 3.6e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 atfg 8
| | | |
DB 98 AFXG 101.

RESULT 17
ID 095792 PRELIMINARY; PRT; 311 AA.
AC 095792.
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CYTOCHROME C (FRAGMENT).
GN CYTB.
OS Emola cyanura (blue-tailed skink).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodonta; Squamata; Scleroglossa; Sclincomorpha; Sclinoidea;
OC Scincidae; Emola.
OX NCB1_TaxID=38244;
RX [1]
RP SEQUENCE FROM N.A.
RA Bruna E.M., Fisher R.N., Case T.J.;
RL Picot, R. Soc. Lond., B, Biol. Sci. 263:681-688(1996).
CC -I- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -I- CATALYTIC ACTIVITY: O₂(2) + 2 FERRICYTOCHROME C = O + 2
FERROCYTOCHROME C.
CC -I- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
BOUND TO THE PROTEIN (BY SIMILARITY).
CC -I- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CYTOCHROME C1 AND THE RISKE PROTEIN (BY SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE CYTOCHROME B/B6 FAMILY.
EMBL; U49347; AAB06601.1; -.
DR InterPro: IPR000179; Cyt_b_b6.

DR Pfam: PF00032; cytochrome_b_c1; 1.
DR Pfam: PF00033; cytochrome_b_n; 1.
DR PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
DR PROSITE: PS00193; CYTOCHROME_B_OO; UNKNOWN.1.
KW Electron transport; Heme; Mitochondrion; Respiratory chain;
FT TRANSMEMBRANE.
FT NON_TER 1
FT NON_TER 311
SQ SEQUENCE 311 AA; 34962 MW; 2CE6D105C1CF0C5 CRC64;

Query Match 37.5%; Score 3; DB 8; Length 311;
Best Local Similarity 100.0%; Pred. No. 3.6e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 atfg 8
| | | |
DB 101 AFXG 104

RESULT 18
ID 092ZG1 PRELIMINARY; PRT; 311 AA.
AC 092ZG1.
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CYTOCHROME B (FRAGMENT).
GN CYTB.
OS Phyllotis osgoodi (Osgood's leaf-eared mouse).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Sigmodontinae;
OC Phyllotis.
OX NCB1_TaxID=59938;
RX [1]
RP SEQUENCE FROM N.A.
RA STRAIN-LCM 638;
RA Steppan S.J.;
RT "Phylogenetic relationships and species limits within Phyllotis
(Rodentia: Sigmodontinae): concordance between mtDNA sequence and
morphology."
RL J. Mammal. 79:0-0(1998).
CC -I- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -I- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
BOUND TO THE PROTEIN (BY SIMILARITY).
CC -I- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CYTOCHROME C1 AND THE RISKE PROTEIN (BY SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE CYTOCHROME B/B6 FAMILY.
EMBL; U86825; AAD12350.1; -.
DR InterPro: IPR000179; Cyt_b_b6.
DR Pfam: PF00032; cytochrome_b_c1; 1.
DR Pfam: PF00033; cytochrome_b_n; 1.
DR PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
DR PROSITE: PS00193; CYTOCHROME_B_OO; UNKNOWN.1.
KW Electron transport; Heme; Mitochondrion; Respiratory chain;
FT TRANSMEMBRANE.
FT NON_TER 1
FT NON_TER 311
SQ SEQUENCE 311 AA; 34708 MW; 3E91294031F728ED CRC64;

Query Match 37.5%; Score 3; DB 8; Length 311;
Best Local Similarity 100.0%; Pred. No. 3.6e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 atfg 8
| | | |
DB 117 AFXG 120

```
RESULT 19
ID 09MRX6 PRELIMINARY; PRT; 326 AA.
AC 09MRX6;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE NADH DEHYDROGENASE (FRAGMENT).
GN NDH.
OS Rhexia mariana.
OC Chloroplast.
CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
CC eustosids II; Myrtales; Melastomataceae; Rhexia.
OX NCBI_TaxID=130523;
RN 11
RP SEQUENCE FROM N.A.
RT Renner S.S., Meyer K.;
RT "Melastomeae come full circle: biogeographic reconstruction and
RT molecular clock dating."
RT Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN 12
RP SEQUENCE FROM N.A.
RT Renner S.S., Meyer K.;
RT Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN 13
RP SEQUENCE FROM N.A.
RT Clausen G., Renner S.S.;
RT Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
CC -I- CATALYTIC ACTIVITY: NADH + PLASTOQUINONE -> NAD(+) + PLASTOQUINOL.
CC -I- SIMILARITY: TO NADH-UBIQUINONE/PLASTOQUINONE (COMPLEX I), VARIOUS
CC CHAINS.
DR EMBL: AF272819; AAF81819.2; -.
DR InterPro: IPR000916; Bel_v_I.
DR InterPro: IPR001750; Oxidored_q1.
DR InterPro: IPR002128; Oxidored_q1_C.
DR Pfam: PF00361; Oxidored_q1; 1.
DR Pfam: PF01010; Oxidored_q1-C; 1.
DR ProDom: PD000531; Bel_v_I; 1.
DR Chloroplast; NAD; Oxidoreductase; Plastocyanine.
KW NON_TER
FT 1
FT NON_TER 326
SQ SEQUENCE 326 AA; 37082 MW; 9B42F1A1B32BB906 CRC64;

Query Match 37.5%; Score 3; DB 8; Length 326;
Best Local Similarity 100.0%; Pred. No. 3.8e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 afxg 8
DB 275 AFXG 278

RESULT 20
ID 003302 PRELIMINARY; PRT; 346 AA.
AC 003302;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CYTOCHROME B (FRAGMENT).
GN CYTB.
OS Campephilus haematogaster.
OC Mitochondrion.
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Archosauria; Aves; Neognathae; Piciformes; Picidae; Campephilus.
OX NCBI_TaxID=56085;
RN 11
RP SEQUENCE FROM N.A.
RT STRAIN=LSU11786;
RT Moore W.S., DeFilippis V.R.;
RT Moore W.S., DeFilippis V.R.;
```

```
RL (In) Mindell D.P. (eds.);
RL Avian molecular evolution and systematics, pp.81-116, Academic Press,
RL New York (1997).
CC -I- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -I- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN (BY SIMILARITY).
CC -I- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE CYTOCHROME B/B6 FAMILY.
DR EMBL: U83284; AAB53600.1; -.
DR InterPro: IPR000179; Cyt_b_b6.
DR Pfam: PF00032; cytochrome_b_c1; 1.
DR Pfam: PF00033; cytochrome_b_n; 1.
DR PROSITE: PS00193; CYTOCHROME_B_QQ; 1.
KW Electron transport; Heme; Mitochondrion; Respiratory chain;
KW Transmembrane.
FT 1
FT NON_TER 346
SQ SEQUENCE 346 AA; 38424 MW; 442E84BF45EEAA2E CRC64;
```

Query Match 37.5%; Score 3; DB 8; Length 346;
Best Local Similarity 100.0%; Pred. No. 4e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 5 afxg 8
DB 94 AFXG 97

RESULT 21
ID 003340 PRELIMINARY; PRT; 347 AA.
AC 003340;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CYTOCHROME B (FRAGMENT).
GN CYTB.
OS Sphyrapicus varius (Yellow-bellied sapsucker).
OC Mitochondrion.
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Archosauria; Aves; Neognathae; Piciformes; Picidae; Sphyrapicus.
OX NCBI_TaxID=56079;
RN 11
RP SEQUENCE FROM N.A.
RT STRAIN=WS086148;
RT Moore W.S., DeFilippis V.R.;
RT (In) Mindell D.P. (eds.);
RT Avian molecular evolution and systematics, pp.81-116, Academic Press,
RL New York (1997).
CC -I- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -I- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN (BY SIMILARITY).
CC -I- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE CYTOCHROME B/B6 FAMILY.
DR EMBL: U83295; AAB53613.1; -.
DR InterPro: IPR000179; Cyt_b_b6.
DR Pfam: PF00032; cytochrome_b_c1; 1.
DR Pfam: PF00033; cytochrome_b_n; 1.
DR PROSITE: PS00193; CYTOCHROME_B_QQ; UNKNOWN_1.
KW Electron transport; Heme; Mitochondrion; Respiratory chain;
KW Transmembrane.
FT 1
FT NON_TER 347
SQ SEQUENCE 347 AA; 38763 MW; D367AF65721BDE0E CRC64;
```

Query Match 37.5%; Score 3; DB 8; Length 347;
Best Local Similarity 100.0%; Pred. No. 4e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 afxg 8
|||||
DB 95 AFXG 98

RESULT 22
ID 003341 PRELIMINARY; PRT; 347 AA.
AC 003341:
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CYTOCHROME B (FRAGMENT).
GN CYTB.
OS Sphyrapicus varius (Yellow-bellied sapsucker).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Piciformes; Picidae; Sphyrapicus.
OX NCBI_TaxID=56079;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSU86149;
RA Moore W.S., Delilippis V.R.;
RL (in) Mindell D.P. (eds.);
RL Avian molecular evolution and systematics, pp.81-116, Academic Press,
RL New York (1997).
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: OX(2) + 2 FERRICYTOCHROME C = O + 2
CC FERROCYTOCHROME C.
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN (BY SIMILARITY).
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B/B6 FAMILY.
CC EMBL: U83296; AAB53614.1; -;
DR InterPro: IPR000179; Cyt_b_b6.
DR Pfam: PF00032; cytochrome_b_c1.1.
DR Pfam: PF00033; cytochrome_b_n.1.
DR PROSITE: PS00192; CYTOCHROME_B_HEME.1.
DR PROSITE: PS00193; CYTOCHROME_B_OO; UNKNOWN.1.
KW Electron transport; Heme; Mitochondrion; Respiratory chain;
KW Transmembrane.
KW NON_TER 1
FT SEQUENCE 347 AA; 38794 MW; 50B8475A78C9E28D CRC64;

Query Match 37.5%; Score 3; DB 8; Length 347;
Best Local Similarity 100.0%; Pred. No. 4e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 afxg 8
|||||
DB 95 AFXG 98

RESULT 23
ID 09TEF8 PRELIMINARY; PRT; 355 AA.
AC 09TEF8:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CYTOCHROME B (FRAGMENT).
GN CYTB.
OS Oryctolagus cuniculus (Rabbit).
OS Mitochondrion.
OC Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euteria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Branco M., Ferrand N., Monnerot M.;
RT *Phylogeography of the European rabbit (Oryctolagus cuniculus) on the
RT Iberian peninsula inferred from RFLP analysis of the cytochrome b
RT gene.";
RT Submitted (JUN-1999) to the EMBL/GenBank/DBS databases
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN (BY SIMILARITY).
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
CC EMBL: AJ243096; CAB45258.1; -;
DR InterPro: IPR000179; Cyt_b_b6.
DR Pfam: PF00032; cytochrome_b_c1.1.
DR Pfam: PF00033; cytochrome_b_n.1.
DR PROSITE: PS00192; CYTOCHROME_B_HEME.1.
DR PROSITE: PS00193; CYTOCHROME_B_OO; UNKNOWN.1.
KW Electron transport; Heme; Mitochondrion; Respiratory chain;
KW Transmembrane.
KW NON_TER 1
FT SEQUENCE 355 AA; 40051 MW; 821F1AF9A85E9 CRC64;

Query Match 37.5%; Score 3; DB 8; Length 355;
Best Local Similarity 100.0%; Pred. No. 4.1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 afxg 8
|||||
DB 103 AFXG 106

RESULT 24
ID 021210 PRELIMINARY; PRT; 374 AA.
AC 021210:
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CYTOCHROME B (FRAGMENT).
GN CYTB.
OS Parachromis dovii.
OS Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Labroidae;
OC Cichlidae; Parachromis.
OX NCBI_TaxID=63177;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=98224076; PubMed=9562979;
RA Martin A.P., Bermingham E.;
RT "Systematics and evolution of lower Central American cichlids inferred
RT from analysis of cytochrome b gene sequences.";
RL Mol. Phylogenet. Evol. 9:192-203(1998).
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN (BY SIMILARITY).
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B/B6 FAMILY.
CC EMBL: AF009924; AAB64130.2; -;

DR InterPro: IPR000179; Cyt_b_b6.
DR Pfam: PF00032; Cytochrome_b_c1.
DR Pfam: PF00033; Cytochrome_b_n; 1.
DR PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
DR PROSITE: PS00193; CYTOCHROME_B_OO; 1.
KW Electron transport; Heme; Mitochondrion; Respiratory chain;
Transmembrane.
KW NON_TER 1
FT NON_TER 374 374
SQ SEQUENCE 374 AA: 41771 MW: 2857DD28DCA24D4 CRC64;

Query Match 37.5%; Score 3; DB 8; Length 374;
Best Local Similarity 100.0%; Pred. No. 4.3e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 afxg 8
|||
DB 125 AFXG 128

RESULT 25
021218

ID 021218 PRELIMINARY; PRT: 379 AA.

DT 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE CYTOCHROME B (FRAGMENT).

CYTB.

Parachromis lolseliei.

OG Mitochondrion.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Acanthopterygii; Perciformes; Labroidel;

OC Cichlidae; Parachromis.

OX NCBI_TaxID=63178;

RN [1]

RP SEQUENCE FROM N.A.

RA Martin A.P., Birmingham E.;

RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.

-1- FUNCTION: COMPONENT OF THE UBINOUL-CYTOCHROME C REDUCTASE

COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A

RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL

COUPLED TO ATP SYNTHESIS (BY SIMILARITY).

-1- CATALYTIC ACTIVITY: OH(2) + 2 FERRICYTOCHROME C + Q + 2

FERROCYTOCHROME C.

-1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY

BOUND TO THE PROTEIN (BY SIMILARITY).

-1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,

CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).

-1- SIMILARITY: BELONGS TO THE CYTOCHROME B/B6 FAMILY.

EMBL: AF009926; AAB64132.1; .

InterPro: IPR000179; Cyt_b_b6.
DR Pfam: PF00032; cytochrome_b_c1.
DR PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
DR PROSITE: PS00193; CYTOCHROME_B_OO; 1.
KW Electron transport; Heme; Mitochondrion; Respiratory chain;
Transmembrane.
KW NON_TER 379 379
FT NON_TER 379 379
SQ SEQUENCE 379 AA: 42399 MW: 2B5895B5481FEBF9 CRC64;

Search completed: January 14, 2002, 08:07:34
Job time: 764 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:56:30 ; Search time 81.39 Seconds
(without alignments)
7.281 Million cell updates/sec

Title: 09-185908-1d
Perfect score: 8
Sequence: 1 wxxxxafgx 8

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size: 0

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 50 summaries

Database:

1: A.Geneseq_1101: *
2: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT: *
3: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT: *
4: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT: *
5: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT: *
6: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT: *
7: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT: *
8: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT: *
9: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT: *
10: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT: *
11: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT: *
12: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT: *
13: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT: *
14: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT: *
15: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT: *
16: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT: *
17: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT: *
18: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT: *
19: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT: *
20: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT: *
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT: *
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37.5	5	19	AAW45592	Substrate #13 for
2	37.5	15	15	AAW51894	Chitin binding pro
3	37.5	23	21	AAV84533	Amino acid sequenc
4	37.5	65	20	AAV36099	Extended human sec
5	37.5	91	22	AAW76045	Human colon cancer
6	37.5	105	18	AAW27978	Amino acid sequenc
7	37.5	112	20	AAV13018	Human secreted pro
8	37.5	127	20	AAV27139	Vertebrate slit pr
9	37.5	162	22	AAW63955	Human prostate can
10	37.5	172	22	AAW73946	Human colon cancer
11	37.5	181	22	AAW80899	Lipid degradation

12	3	37.5	181	22	AAW83349	P patens lipid met
13	3	37.5	186	22	AAW43545	Human polyPeptide
14	3	37.5	392	22	AAE04176	Human gene 2 encod
15	3	37.5	414	22	AAE04125	Human gene 2 encod
16	3	37.5	503	21	AAW63190	Human secreted pro
17	2	25.0	3	15	AAW42556	ACE inhibitor sp3.
18	2	25.0	3	15	AAW85059	Calcium ion channe
19	2	25.0	3	19	AAW76946	Fusion Immunoglob
20	2	25.0	3	20	AAW84197	Peptide comprising
21	2	25.0	3	21	AAV49387	Dermorphin derivat
22	2	25.0	4	2	AAW10545	Analgesic peptide
23	2	25.0	4	2	AAW10584	Analgesic peptide
24	2	25.0	4	2	AAW10547	Analgesic peptide
25	2	25.0	4	2	AAW10548	Analgesic peptide
26	2	25.0	4	2	AAW10550	Analgesic peptide
27	2	25.0	4	2	AAW10262	Narcotic agonist p
28	2	25.0	4	2	AAW10264	Narcotic agonist p
29	2	25.0	4	2	AAW10265	Narcotic agonist p
30	2	25.0	4	2	AAW10266	Narcotic agonist p
31	2	25.0	4	2	AAW10267	Narcotic agonist p
32	2	25.0	4	2	AAW10268	Narcotic agonist p
33	2	25.0	4	2	AAW10269	Narcotic agonist p
34	2	25.0	4	2	AAW10270	Narcotic agonist p
35	2	25.0	4	2	AAW10260	Peptide alcohol 2.
36	2	25.0	4	3	AAW20058	Enzyme substrate s
37	2	25.0	4	5	AAW40746	Dermorphin deriva
38	2	25.0	4	5	AAW40410	Chromogenic serine
39	2	25.0	4	6	AAW50696	Sequence of Pharma
40	2	25.0	4	6	AAW50545	Sequence of Cyclo
41	2	25.0	4	7	AAW60801	Peptide with neut
42	2	25.0	4	9	AAW81591	Organ specific neo
43	2	25.0	4	10	AAW91623	Motif useful in to
44	2	25.0	4	12	AAW13033	Isoenin deriv. syn
45	2	25.0	4	13	AAW20049	Tetrapeptide chole
46	2	25.0	4	13	AAW25367	Peptide #8 pred a
47	2	25.0	4	13	AAW30195	Des-Tyr(1)-beta-ca
48	2	25.0	4	13	AAW30202	Des-Tyr(1)-beta-ca
49	2	25.0	4	13	AAW30206	Des-Tyr(1)-beta-ca
50	2	25.0	4	14	AAW41644	Internalisation si

ALIGNMENTS

RESULT 1	
AAW45592	standard; peptide; 5 AA.
ID	
XX	
AC	AAW45592:
DT	
DT	04-JUN-1998 (first entry)
XX	
DE	Substrate #13 for the determination of pepsin enzymic activity.
XX	
KW	Protease activity; pepsin; pepsinogen; substrate; thiol derivative.
XX	
OS	Synthetic.
XX	
FT	Key
FT	Modified-site 1 Location/Qualifiers
FT	Modified-site 4 /note= "benzyloxycarbonyl-Ala"
FT	Modified-site 5 /note= "beta-phenylthiolactic acid"
FT	Modified-site 5 /note= "Gly-4-amino pyridine"
PN	JP09295997-A.
XX	
PD	18-NOV-1997.
XX	
PF	07-MAY-1996: 96JP-0112441.
XX	
PR	07-MAY-1996: 96JP-0112441.

XX	(NITTO) NITTO BOSEKI CO LTD.
PA	WPI; 1998-046962/05.
DR	
XX	
XX	New substrate for the determination of pepsin enzymatic activity -
PT	by allowing the enzyme to hydrolyse the substrate and then measuring
PT	the amount of thiol derivative formed
XX	
PS	Example 13; Page 8; 12pp; Japanese.
XX	
CC	The present sequence represents a peptide substrate for the
CC	determination of pepsin activity. The invention relates to peptides
CC	of the following formula: X-A-B-C (1); X = H, a group masking a terminal
CC	amino group irreversibly or a protective group for an amino group used
CC	usually in peptide chemistry; A = glycine, D- or L-alanine, valine,
CC	norvaline, leucine, isoleucine, norleucine, serine and threonine, or a
CC	dipeptide residue consisting of two of these amino acids; B =
CC	phenylalanine, parantitrophenylalanine or tyrosine; C = thiol derivative
CC	residue constituted by 1 to 30 C atoms which can form thioester bond and
CC	H atoms combining with them in which 1 to 10 C atoms can be substituted
CC	by O, N and/or S and 2 to 10 H atoms can be substituted by O, N and/or
CC	S, or its salt. The peptides can be used to determine the
CC	presence/activity of pepsin or pepsinogen in a sample by determining the
CC	amount of thiol derivative formed due to the action (ie. hydrolysis) of
CC	the enzyme on the peptides.
XX	
SO	Sequence 5 AA;
Query Match	37.5%; Score 3; DB 19; Length 5;
Best Local Similarity	100.0%; Pred. No. 4,3e+05;
Matches 4;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	5 atfg 8
DB	2 atfg 5
RESULT 2	
AAAS1894	
ID	AAAS1894 standard; peptide: 15 AA.
XX	
AC	AAAS1894;
XX	
DT	09-SEP-1994 (first entry)
XX	
DE	Chitin binding protein internal peptide.
XX	
KW	Chitin binding protein; CBP; antifungal; fungicide;
KW	beta-1,3-glucanase; transgenic plant; disease-resistance;
KW	crop improvement; tobacco.
XX	
OS	Nicotiana tabacum.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 5 /note= "Probably tyrosine"
FT	11 /note= "probably proline or arginine"
FT	Misc-difference 11 /note= "probably proline or arginine"
XX	
PN	W09408009-A.
XX	
PD	14-APR-1994.
XX	
PE	05-OCT-1993; 93MO-EP02790.
XX	
PR	05-OCT-1992; 92EP-0203071.
PR	13-MAY-1993; 93EP-0201370.
XX	
PA	(MOGE-) MOGEN INT NV.
XX	
XX	Apotheker M, Bres-vloemans AA, Cornelissen BJC, Melchers LS;

PI	Ponstein AS, Sela-buurlage MB;
XX	
DR	WPI, 1994-135576/16.
XX	
PT	New antifungal chitin binding protein from plants - without
PT	significant chitinase activity, showing synergistic activity with
PT	1,3-beta-glucanase, also its nucleic acid, vectors, transformed
PT	plants, etc.
XX	
PS	Disclosure; Page 31; 54pp; English.
XX	
CC	Internal peptides of tobacco cv. Samsun NN chitin binding protein
CC	were obtained by digestion with protease V8 (peptide AAR51895) and N-
CC	chlorosuccinimide/urea (peptide AAR51894). CBP can be expressed in
CC	transgenic plants to improve resistance to fungal pathogens.
SQ	Sequence 15 AA;
QY	5 atxg 8
DB	3 atfxg 6
RESULT 3	
AAV84533	37.5%; Score 3; DB 15; Length 15;
ID AAV84533 standard; peptide: 23 AA.	Best Local Similarity 100.0%; Pred. No. 3.3e+02;
XX	
AC AAV84533;	Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX	
DT 25-JUL-2000 (first entry)	
XX	
XX	Amino acid sequence of a modified enterotoxin II toxin signal peptide.
DE	
XX	Enterotoxin II; signal peptide; peptide secretion; human growth hormone.
KW	
XX	Synthetic.
OS	Escherichia coll.
XX	
FT	Key Location/Qualifiers
FT	Misc-difference 4
FT	/label= Ser, Thr, Lys, Gln
FT	Misc-difference 5
FT	/label= Ile, Thr, Ser
FT	Misc-difference 12
FT	/label= Ala, Gly, Val, Leu, Ile
FT	Misc-difference 20
FT	/label= Ile, Phe, Ala, Val
FT	Misc-difference 22
FT	/label= Gln, Asn, Ala, Lys
XX	
PN	WO200015661-A1.
XX	
PD	23-MAR-2000.
XX	
PF	15-SEP-1999; 99WO-KR00547.
XX	
PR	15-SEP-1998; 98KR-0038061.
XX	
PA	(HANMI-) HANMI PHARM CO LTD.
XX	
PI	Kwon SC, Jung SY, Shin H, Choi JD, Choi KD, Lee GS;
XX	
DR	WPI; 2000-271381/23.
XX	
PT	Modified Escherichia coli enterotoxin II signal peptides useful for
PT	enhancing protein, especially growth hormone, secretion from
PT	recombinant cells -
XX	

PS Clafim 2; Page -: 43pp; English.

XX AAY84533-35 represent modified Escherichia coli enterotoxin II

CC signal peptides. The specification describes modified enterotoxin II

CC signal peptides, which are involved in directing the secretion of

CC enterotoxin II from the cell. The modified peptides are characterized

CC in that at least one of the 2nd, 4th, 5th, 12th, 20th and 22nd amino

CC acids of the enterotoxin II signal peptide is replaced by another amino

CC acid, and at least 1 of the 2nd and 4th amino acids of the modified

CC peptide is lysine. The modified signal peptide enhances the efficiency

CC of peptide secretion from the cells. The modified signal peptides may

CC be used according to standard recombinant DNA methodologies to direct

CC the secretion of peptides from microorganisms. In particular, it may

CC be used to direct the secretion of human growth hormone from E. coli

CC cells in fermentation culture.

CC note: this sequence does not appear in the specification; it was created

CC using information provided.

XX Sequence 23 AA:

QY 37.5%; Score 3; DB 21; Length 23;

Best Local Similarity 100.0%; Pred. No. 4.3e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 kxxaf 6

DB 3 kxxaf 7

RESULT 4

AAV36099 standard; Protein; 65 AA.

XX AAY36099;

AC AAY36099;

XX 13-SEP-1999 (first entry)

DT 13-SEP-1999 (first entry)

DE Extended human secreted protein sequence, SEQ ID NO. 484.

XX Secreted protein; human; cytokine; cellular proliferation; cell movement;

KW cellular differentiation; immune system regulator; anti-inflammatory;

KW haematopoiesis regulator; tissue growth regulator; tumour inhibitor;

KW reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;

KW genetic disease.

XX Homo sapiens.

OS Homo sapiens.

XX WO9931236-A2.

PN 24-JUN-1999.

XX 17-DEC-1998; 98WO-1B02122.

PF 10-AUG-1998; 98US-0096116.

PR 17-DEC-1997; 97US-0069957.

PR 09-FEB-1998; 98US-0074121.

PR 13-APR-1998; 98US-0081563.

XX (GEST) GENSET.

XX Bougueleret L, Duclet A, Dumas Milne Edwards J;

PI MPI; 1999-385906/32.

DR N-PSDB; AAX97783.

XX New isolated human secreted proteins

PT Clalm 9; Page 419; 516pp; English.

PS This sequence is encoded by an extended human secreted protein coding

CC sequence of the invention. The secreted proteins can be used in treating

CC or controlling a variety of human conditions. The secreted proteins may

CC act as cytokines or may affect cellular proliferation or differentiation

CC or may act as immune system regulators, haematopoiesis regulators, tissue

CC growth regulators, regulators of reproductive hormones or cell movement

CC or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or

CC tumour inhibition activity. The DNAs can be used in forensic procedures

CC to identify individuals or in diagnostic procedures to identify

CC individuals having genetic diseases resulting from abnormal expression of

CC the genes corresponding to the extended cDNAs. They are also useful for

CC constructing a high resolution map of the human chromosomes. They can

CC also be used for gene therapy to control or treat genetic diseases.

XX Sequence 65 AA:

QY 37.5%; Score 3; DB 20; Length 65;

Best Local Similarity 100.0%; Pred. No. 8e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 afxg 8

DB 55 afxg 58

RESULT 5

AAG76045 standard; Protein; 91 AA.

XX AAG76045;

AC AAG76045;

XX 03-SEP-2001 (first entry)

DT 03-SEP-2001 (first entry)

DE Human colon cancer antigen protein SEQ ID NO:6809.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;

KW colorectal carcinoma; chromosome 2.

XX Homo sapiens.

OS Homo sapiens.

XX WO200122920-A2.

PN 05-APR-2001.

PD 28-SEP-2000; 2000WO-US26524.

PE 29-SEP-1999; 99US-0157137.

PR 03-NOV-1999; 99US-0163280.

PR (HUMA-) HUMAN GENOME SCI INC.

PA Ruben SM, Barash SC, Birse CE, Rosen CA;

PI MPI; 2001-235357/24.

DR N-PSDB; AAH35450.

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,

PT useful for preventing, diagnosing and/or treating colorectal cancers -

PS Claim 11; Page 8261-8262; 9803pp; English.

XX AAH32943 to AAH37195 and AAG73514 to AAG77768 represent human colon

CC cancer-associated nucleic acid molecules (N) and proteins (P), where

CC the proteins are collectively known as colon cancer antigens. The colon

CC cancer antigens have cytostatic activity and can be used in gene

CC therapy and vaccine production. N and P may be used in the prevention,

CC diagnosis and treatment of diseases associated with inappropriate P

CC expression. For example, N and P may be used to treat disorders

CC associated with decreased expression by rectifying mutations or deletions

CC in a patient's genome that affect the activity of P by expressing

CC inactive proteins or to supplement the patient's own production of P.

CC Additionally, N may be used to produce the colon cancer-associated P,

CC by inserting the nucleic acids into a host cell and culturing the cell

CC to express the proteins. N and P can be used in the prevention, diagnosis

CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204

CC and AAB77789 represent sequences used in the exemplification of the
CC present invention. CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

SO Sequence 91 AA;

Query Match 37.5%; Score 3; DB 22; Length 91;
Best Local Similarity 100.0%; Pred. No. 9.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 5 atxg 8
| | | |
Db 83 atxg 86

RESULT 6

AAM27978
ID AAM27978 standard; Protein: 105 AA.

AC AAM27978;

DE 26-AUG-1998 (first entry)

XX Amino acid sequence of branched amino acid transport system protein.

XX Staphylococcus aureus protein; ribozyme; antisense sequence; control;

KW Staphylococcal gene; regulatory element; bacterial gene expression;

KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;

XX toxic shock syndrome.

OS Staphylococcus aureus.

XX Key Location/Qualifiers

FT Misc-difference 1..105

XX "residues designated X are not defined in

XX the specification"

XX WO9730070-A1.

XX 21-AUG-1997.

XX 19-FEB-1997; 97MO-US02318.

XX 20-FEB-1996; 96US-0011888.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO;

XX Pratt JM, Reichard RW, Rosenberg M, Ward JM;

XX WPI: 1997-424969/39.

XX N-PSDB; AAT83937.

XX Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used

XX to isolate antimicrobial compounds, and in vaccines against S.

XX aureus infection

PS Claim 6; Page 385-386; 9899pp; English.

XX The present sequence represents a Staphylococcus aureus protein, that,

XX based on homology is believed to be a membrane-associated component of a

XX branched amino acid transport system. The DNA sequence was isolated from

XX a library of clones of S. aureus WCUH 29 in Escherichia coli. The DNA

XX sequence can be used in the construction of ribozymes and antisense

XX sequences to control the expression of Staphylococcal genes. The DNA

XX sequence is also useful as a source of regulatory elements for the

XX control of bacterial gene expression. The present protein may be used

XX to produce vaccines to enable a host to produce specific antibodies

XX with antibacterial action. These vaccines and antibodies would protect

XX a host,against invasion by S. aureus, and conditions relating to

CC Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled

CC skin syndrome, and toxic shock syndrome.

XX SQ Sequence 105 AA;

Query Match 37.5%; Score 3; DB 18; Length 105;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 5 atxg 8
| | | |
Db 47 atxg 50

RESULT 7

AAV13018
ID AAV13018 standard; Protein: 112 AA.

AC AAV13018;

DE 22-JUN-1999 (first entry)

XX Human secreted protein encoded by 5' EST SEQ ID NO: 32.

XX Human; secreted protein; EST; expressed sequence tag; diagnosis;

KW forensic; gene therapy; chromosome mapping; signal peptide;

KW upstream regulatory sequence; cytokine activity; cell proliferation;

KW differentiation; haematopoiesis regulation; tissue growth regulation;

KW reproductive hormone regulation; chemotactic; chemokine; haemostatic;

KW thrombolytic; anti-inflammatory; tumour inhibition.

XX Homo sapiens.

OS WO9906552-A2.

XX 11-FEB-1999.

XX 31-JUL-1998; 98MO-IB01236.

XX 01-AUG-1997; 97US-0905223.

XX (GENSET) GENSET.

XX Duclert A, Dumas Milne Edwards J, Lacroix B;

XX WPI: 1999-153782/13.

XX N-PSDB; AAX51818.

XX New isolated brain-derived nucleic acids - used to develop products

XX which may have cytokine, immune, regulatory, haematopoiesis

XX regulating, anti-inflammatory or tumour inhibition activity

PS Claim 34; Page 454; 577pp; English.

XX AAX51787 to AAX52019 represent 5' expressed sequence tags (ESTs) for

XX human secreted proteins, and encode the proteins given in AAY12987 to

XX AAY13219, respectively. The proteins given represent the signal peptide

XX and an N-terminal fragment of a secreted protein. The nucleic acid

XX sequences can be used for producing secreted human gene products. They

XX can also be used to develop products for diagnosis and therapy. The

XX proliferation/differentiation activity, haematopoiesis regulating

XX activity, tissue growth regulating activity, reproductive hormone

XX regulating activity, chemotactic/chemokine activity, haemostatic and

XX thrombolytic activity, receptor/ligand activity, anti-inflammatory

XX activity, tumour inhibition activity or other activities. The products

XX can be used in forensic, gene therapy and chromosome mapping procedures.

XX The sequences can also be used for obtaining corresponding promoter

XX sequences. The nucleic acids encoding the signal peptide can be used for

XX directing extracellular secretion of a polypeptide or the insertion of a

XX polypeptide into a membrane, or importing a polypeptide into a cell.

SO Sequence 112 AA:

Query Match 37.5%; Score 3; DB 20; Length 112;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 afxg 8
||||
Db 64 atxg 67

RESULT 8

AAV27139
ID AAV27139 standard; protein; 127 AA.

XX AAV27139;

DT 15-SEP-1999 (first entry)

DE Vertebrate slit protein (Seq ID No: 3 of JP11164690).

KW Vertebrate-derived protein; slit protein; diagnosis; cancer; nerve;
KW muscle; endocrine system.

XX Vertebra.

OS Vertebra.

XX Key Location/Qualifiers
FH Misc-difference 1..150
FT /note: "Residues Xaa can be any amino acid"

PN JP11164690-A.

PD 22-JUN-1999.

PF 05-DEC-1997; 97JP-0335435.

PR 05-DEC-1997; 97JP-0335435.

PA (ASAH) ASAMI KASEI KOGYO KK.

DR WPI: 1999-411830/35.

XX New vertebrate slit protein - useful for diagnosis and treatment of
PT cancers in nerves; muscle and endocrine system

PS Claim 1; Page 38-39; 102pp; Japanese.

XX The invention relates to a vertebrate-derived protein containing an

CC amino acid sequence shown in AAV27137 and AAV27139. The

CC vertebrate-derived protein has at least 55 % homology to one of sequences

CC shown in AAV27141-Y27143, and has slit protein-like activity. The

CC vertebrate slit proteins encoding nucleic acid sequences have at least

CC 60% homology to nucleic acid sequences AAX89161-163. The

CC vertebrate-derived proteins can be produced recombinantly by transforming

CC host cells with expression vectors comprising the encoding nucleic acids.

CC The proteins of the invention are for diagnosing and treating cancer of

CC the nerves, muscle and/or endocrine system.

XX Sequence 127 AA:

Query Match 37.5%; Score 3; DB 20; Length 127;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 afxg 8
||||
Db 40 atxg 43

RESULT 9
AAB63955

ID AAB63955 standard; Protein; 162 AA.

XX AAB63955;

DT 26-MAR-2001 (first entry)

DE Human prostate cancer associated antigen protein sequence SEQ ID NO:1317.

KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW cancer associated antigen; cytostatic; cancer vaccine.

XX Homo sapiens.

OS Homo sapiens.

PN WO200073801-A2.

PD 07-DEC-2000.

PF 26-MAY-2000; 2000WO-US14749.

PR 28-MAY-1999; 99US-0136526.

PR 10-SEP-1999; 99US-0153454.

PA (LUDW-) LUDWIG INST CANCER RES.

PI Odata Y;

DR WPI: 2001-025274/03.

XX Nucleic acids encoding breast, gastric and prostate cancer associated

PT antigen precursors, useful for diagnosing and treating a condition

PT characterized by expression of an abnormal amount of a protein, e.g.

PT cancer -

XX Example 1; Page 792; 799pp; English.

XX AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014

CC represent nucleotide sequences encoding human breast, gastric and

CC prostate cancer associated antigen precursors (CAAP) respectively.

CC AAB6332 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970

CC represent human breast, gastric and prostate CAAP protein sequence

CC respectively; CAAPs have cytostatic activity and can be used in the

CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic

CC acids or anti-CAAP antibodies are useful for diagnosing and treating a

CC condition characterized by expression of an abnormal amount of a protein,

XX e.g. cancer.

XX Sequence 162 AA;

Query Match 37.5%; Score 3; DB 22; Length 162;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 afxg 8
||||
Db 128 atxg 131

RESULT 10

AAG73946
ID AAG73946 standard; Protein; 172 AA.

XX AAG73946;

DT 03-SEP-2001 (first entry)

DE Human colon cancer antigen protein SEQ ID NO:4710.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 11.

XX Homo sapiens.

PA (BABI) BASF PLANT SCI GMBH.
XX
PI Lerchl J, Renz A, Ehrhardt T, Reindl A, Cirpus P, Bischoff F;
PI Frank M, Freund A, Duvenig E, Schmidt R, Reski R;
XX WPI: 2001-381293/40.
XX
PT New isolated nucleic acid molecule encoding lipid Metabolism Related
XX Proteins useful in the production of fine chemicals -
XX
PS Clalm 31; Page 106; 113pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of moss lipid metabolism related proteins (LMRPs). The moss
CC Physcomitrella patens is one of the few plants able to produce
CC polyunsaturated fatty acids, and the sequences can be used to create
CC transgenic plants also capable of producing them. They can also be used
CC to identify the presence of P. patens and in the production of fine
CC chemicals. The present sequence is one of the proteins of the invention.
XX
SQ Sequence 181 AA:

Query Match 37.5%; Score 3; DB 22; Length 181;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 afxg 8
|||
Db 5 afxg 8

RESULT 13
AAM43545
ID AAM43545 standard; Protein; 186 AA.
XX
AC AAM43545;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 223.
XX
KW Human; arthritic; antirheumatic; antiproliferative; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; antitumor; antiviral; anticonvulsant; antibacterial;
KW antiparasitic; cardiac; gene therapy; cancer; immune disorder;
KW cardiovascular disorder; neurological disease; infection; human.
XX
OS Homo sapiens.
XX
PN WO200155308-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01309.
XX
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216447.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239955.
PR 13-OCT-2000; 2000US-0239957.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-488781/53.
DR N-PSDB; AAI63851.
XX
XX New isolated nucleic acids and polypeptides, useful for diagnosing,
PT treating and/or preventing human diseases and disorders -
XX
XX Claim 11; SEQ ID NO 223; 664pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AAI63803-AAI64012) and
CC the encoded proteins (AAM434497-AAM43660) useful for preventing, treating
CC or ameliorating medical conditions e.g. by protein or gene therapy. The
CC genes were isolated from a range of human tissues disclosed in the
CC specification. The nucleic acids, proteins, antibodies and (ant)agonists
CC are useful in the diagnosis, treatment and prevention of: (a) cancer,
CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,
CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or
CC uterine; (b) immune disorders e.g. Addison's disease, allergies, mel-
CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
XX

CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pft_sequences.
XX
SQ Sequence 186 AA:

Query Match 37.5%; Score 3; DB 22; Length 186;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 afxg 8
Db 73 afxg 76

RESULT 14
ID AAE04176 standard; Protein; 392 AA.
XX
AC AAE04176;
XX
DT 09-AUG-2001 (first entry)
XX
DE Human gene 2 encoded secreted protein fragment. SEQ ID NO:167.
XX
XX Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
KW psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
KW inflammation; neurological disorder; Alzheimer's disease; food additive;
KW angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
KW pregnancy-related disorder; endocrine disorder; infection; wound healing;
KW cell culture; chemotaxis; vulnerability; binding partner identification;
KW gene therapy.
XX
XX Homo sapiens.
OS
FH Key Location/Qualifiers
FT Misc-difference 152 /label= Unknown
FT /note= "xaa equals any of the naturally occurring
FT L-amino acids"
FT Misc-difference 388 /label= Unknown
FT /note= "xaa equals any of the naturally occurring
FT L-amino acids"
XX
XX WO200134643-A1.
XX
XX 17-MAY-2001.
XX
XX 08-NOV-2000; 2000WO-US30629.
XX
XX 12-NOV-1999; 99US-0164825.
XX 03-AUG-2000; 2000US-0222904.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis GA, Soppet DR, Shi Y;
XX
XX WPI; 2001-374441/39.
XX
XX Nucleic acids encoding 24 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. Gaucher's disease,
PT Alzheimer's disease, Schmitz syndrome, Creutzfeldt-Jacob disease,
PT diabetes mellitus and multiple sclerosis -
XX
PS Disclosure; Page 10; 532pp; English.

XX AAD08404-AAD08478 represent cDNAs corresponding to 24 human secreted
CC protein genes, and AAE04100-AAE004170 represent the proteins they encode.
CC AAE04172-AAE04197 represent human secreted protein fragments or variants.
CC The secreted proteins and their genes are useful for preventing, treating,
CC or ameliorating medical conditions, e.g., by protein or gene therapy.
CC Pathological conditions can be diagnosed by determining the amount of the
CC new protein in a sample or by determining the presence of mutations in
CC the new genes. Specific uses are described for each of the 24 genes,
CC based on the tissues in which they are most highly expressed, and include
CC developing products for the diagnosis or treatment of proliferative
CC disorders, cancer, tumours, foetal and developmental abnormalities,
CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,
CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
CC angiogenic disorders, kidney disorders, gastrointestinal disorders,
CC pregnancy-related disorders, endocrine disorders, and infections. The
CC proteins can also be used to aid wound healing and epithelial cell
CC proliferation, to prevent skin aging due to sunburn, to maintain organs
CC before transplantation, for supporting cell culture of primary tissues,
CC to regenerate tissues, to identify their cognate ligands or binding
CC partners, and in chemotaxis, and can be used as a food additive or a
CC preservative to modify storage properties. Antibodies specific for a
CC protein of the invention can be used in alleviating symptoms associated
CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,
CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The
CC present sequence represents a human secreted protein fragment referred
CC to in the disclosure of the invention.

SO Sequence 392 AA:

Query Match 37.5%; Score 3; DB 22; Length 392;

Best Local Similarity 100.0%; Pred. No. 2.4e+03; Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 afxg 8
|||||
Db 386 afxg 389

RESULT 15

ID AAE04125 standard; Protein; 414 AA.

XX AAE04125;

AC 09-AUG-2001 (first entry)

XX Human gene 2 encoded secreted protein HMLFE89, SEQ ID NO:112.

XX Human: secreted protein; proliferative disorder; cancer; tumour; asthma;
KW focal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
KW psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
KW inflammation; neurological disorder; Alzheimer's disease; food additive;
KW angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
KW pregnancy-related disorder; endocrine disorder; infection; wound healing;
KW cell culture; chemotaxis; vulnerability; binding partner identification;
KW gene therapy.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..26 /label= signal_peptide

FT Protein 27..414 /label= Mature_human_secreted_protein

FT Misc-difference 174 /label= Unknown

FT /note= "Encoded by GKT"

FT Misc-difference 410

FT /label= Unknown

FT /note= "Encoded by CAN"

XX WO200134643-A1.

XX 17-MAY-2001.

XX 08-NOV-2000; 2000WO-US30629.

XX 12-NOV-1999; 99US-0164825.

XX 03-AUG-2000; 2000US-0222904.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Komatsoulis GA, Soppet DR, Shi Y;

XX WPI: 2001-374441/39.

XX N-PSDB: AAD08430.

XX Claim 11: Page 480-482; 532pp; English.

CC AAD08404-AAD08478 represent cDNAs corresponding to 24 human secreted
CC protein genes, and AAE04100-AAE004170 represent the proteins they encode.
CC AAE04172-AAE04197 represent human secreted protein fragments or variants.
CC The secreted proteins and their genes are useful for preventing, treating,
CC or ameliorating medical conditions, e.g., by protein or gene therapy.
CC Pathological conditions can be diagnosed by determining the amount of the
CC new protein in a sample or by determining the presence of mutations in
CC the new genes. Specific uses are described for each of the 24 genes,
CC based on the tissues in which they are most highly expressed, and include
CC developing products for the diagnosis or treatment of proliferative
CC disorders, cancer, tumours, foetal and developmental abnormalities,
CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,
CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
CC angiogenic disorders, kidney disorders, gastrointestinal disorders,
CC pregnancy-related disorders, endocrine disorders, and infections. The
CC proteins can also be used to aid wound healing and epithelial cell
CC proliferation, to prevent skin aging due to sunburn, to maintain organs
CC before transplantation, for supporting cell culture of primary tissues,
CC to regenerate tissues, to identify their cognate ligands or binding
CC partners, and in chemotaxis, and can be used as a food additive or a
CC preservative to modify storage properties. Antibodies specific for a
CC protein of the invention can be used in alleviating symptoms associated
CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,
CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA).
CC The present sequence represents a human secreted protein of
CC the invention.

SO Sequence 414 AA;

Query Match 37.5%; Score 3; DB 22; Length 414;

Best Local Similarity 100.0%; Pred. No. 2.4e+03; Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 afxg 8
|||||
Db 408 afxg 411

RESULT 16

ID AAB63190 standard; Protein; 503 AA.

XX AAB63190;
 XX
 XX 26-MAR-2001 (first entry)
 XX
 DE Human secreted protein sequence encoded by gene 5 SEQ ID NO:116.
 XX
 KM Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
 KM antirheumatic; antiproliferative; cytostatic; cariant; vasotropic;
 KM cerebroprotective; neurotropic; neuroprotective; antibacterial; vitruclide;
 KM fungicide; ophthalmological; gene therapy; autoimmune disease; neoplasm;
 KM rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
 KM cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
 KM angiogenesis; nervous system disorder; Alzheimer's disease; infection;
 KM ocular disorder; corneal infection; wound healing; skin aging;
 KM food additive; preservative.
 XX
 OS Homo sapiens.
 XX
 PN W0200061629-A1.
 XX
 PD 19-OCT-2000.
 XX
 PF 06-APR-2000; 2000WO-US09071.
 XX
 PR 09-APR-1999; 99US-0128694.
 PR 20-JAN-2000; 2000US-0176931.
 XX
 PA (HOMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 XX
 PI Ruben SM, Komatsoulis G;
 PI
 DN WPI; 2000-647420/62.
 XX
 PT Isolated nucleic acid molecule encoding a human secreted protein is
 PT used in preventing, treating or ameliorating a medical condition -
 XX
 PS Disclosure: Page 489-491; 533pp; English.
 CC AAF22373 to AAF22421 encode the human secreted proteins given in AAB63134
 CC to AAB63182. AAB63183 to AAB63231 represent more human secreted proteins
 CC and polypeptides homologous to them. Human secreted proteins have
 CC activities based on the tissues and cells the genes are expressed in.
 CC Examples of activities include: immunosuppressive; antiarthritic;
 CC antirheumatic; antiproliferative; cytostatic; cariant; vasotropic;
 CC cerebroprotective; neurotropic; neuroprotective; antibacterial; vitruclide;
 CC fungicide; and ophthalmological. The polynucleotides and proteins can be
 CC used to prevent, treat or ameliorate a medical condition in e.g. humans,
 CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are
 CC also used in diagnosing a pathological condition or susceptibility to a
 CC pathological condition. Disorders which are diagnosed or treated include
 CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular
 CC disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral
 CC ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's
 CC disease, infections caused by bacteria, viruses and fungi and ocular
 CC disorders e.g. corneal infection. The polypeptides can also be used to
 CC aid wound healing and epithelial cell proliferation, to prevent skin
 CC aging due to sunburn, to maintain organs before transplantation, for
 CC supporting cell culture of primary tissues, to regenerate tissues and in
 CC chemotaxis. The polypeptides can also be used as a food additive or
 CC preservative to increase or decrease storage capabilities. AAF22364 to
 CC AAF22372 and AAB63133 represent sequences used in the exemplification of
 CC the present invention.
 CC
 XX
 SQ Sequence 503 AA;

Query Match 37.5%; Score 3; DB 21; Length 503;
 Best Local Similarity 100.0%; Pred. No. 2.7e+03;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 afxg 8
 ||||
 Db 49 afxg 52

RESULT 17
 AAR42556
 ID AAR42556 standard; peptide; 3 AA.
 XX
 AC AAR42556;
 XX
 DT 07-DEC-1994 (first entry)
 DE ACE inhibitor SP3.
 XX
 KW Angiotensin converting enzyme; ACE; inhibitor; SP5; SP4; SP3;
 KW hypertension; blood pressure.
 XX
 OS Synthetic.
 XX
 PN JP06065288-A.
 XX
 PD 08-MAR-1994.
 XX
 PF 19-AUG-1992; 92JP-0220270.
 XX
 PR 19-AUG-1992; 92JP-0220270.
 XX
 PA (APIA-) API KK.
 PA
 DR WPI; 1994-115194/14.
 XX
 PT New tri-, tetra- and penta-peptide(s), e.g. Trp-Lys-Tyr - are ACE
 PT inhibitors useful for treatment or prophylaxis of hypertension
 XX
 PS Claim 1; Page 2; 5pp; Japanese.
 XX
 CC Peptides SP3, SP4 and SP5 have ACE inhibiting activity. They can be
 CC prepd. easily and in high yield. They are useful for treatment or
 CC prophylaxis of hypertension.
 CC
 XX
 SQ Sequence 3 AA;

Query Match 25.0%; Score 2; DB 15; Length 3;
 Best Local Similarity 100.0%; Pred. No. 4.3e+05;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
 ||
 Db 1 wk 2

RESULT 18
 AAR85069
 ID AAR85069 standard; peptide; 3 AA.
 XX
 AC AAR85069;
 XX
 DT 13-JUN-1996 (first entry)
 DE Calcium ion channel blocking peptide fragment #2.
 DE
 XX Calcium ion channel; blocking peptide; hypotensive agent; blood pressure;
 KW therapy.
 KW
 XX
 OS Synthetic.
 XX
 PN JP07278185-A.
 XX
 PD 24-OCT-1995.
 XX
 PR 08-APR-1994; 94JP-0070579.

XX 08-APR-1994; 94JP-0070579.
PR (FARRH) HOECHST JAPAN KK.
PA WPI: 1995-400979/51.
DR Peptide having calcium channel blocking activity - useful in
PT hypotensive agent
XX Claim 2: Page 2; 6pp; Japanese.
PS AAR85068 and AAR85069 represent calcium ion channel blocking peptides.
CC This sequence corresponds to residues 10-13 of the peptide represented
CC by AAR85068. The peptides suppress blood pressure by inhibiting the
CC calcium influx into cells. These peptides can be used in a hypotensive
CC agent.
XX Sequence 3 AA:
SQ

Query Match 25.0%; Score 2; DB 16; Length 3;
Best Local Similarity 100.0%; Pred. No. 4.3e+05; Mismatches 0; Indels 0; Gaps 0;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
II
DB 2 af 3

RESULT 19
AAW76946
ID AAW76946 standard; peptide; 3 AA.
XX
AC AAW76946;
XX
DT 25-JAN-1999 (first entry)
XX
DE Fusion Immunoglobulin heavy chain HIV gp120 B cell epitope #86.
XX
KW B cell; T cell; epitope; immunoglobulin; heavy chain; gp120; IGH;
KW human immune deficiency virus; HIV; tolerance; treatment; therapy;
KW prophylaxis; vaccine; chemotherapy; immune response; modifier; tumour;
KW microbial infection; autoimmune disease; antibody; apoptosis;
KW antiviral T cell immunity.
XX
OS Mus sp.
OS Homo sapiens.
XX
XX WO9836087-A1.
XX
PD 20-AUG-1998.
XX
XX 13-FEB-1998; 98WO-US02766.
XX
PR 13-FEB-1997; 97US-0040581.
XX
PA (AMNA-) AMERICAN NAT RED CROSS.
XX
PI Scott D, Zambidis E;
XX
XX WPI: 1998-506315/43.
XX
DR New fusion immunoglobulin heavy chain including gp120 epitopes and
XX related complete antibodies - DNA, vectors and transformed cells,
XX used to induce tolerance to the epitopes for treatment of human
XX Immune deficiency virus infection
XX
XX Disclosure; Page 40; 154pp; English.
XX
XX This sequence is an epitope used in the construction of a novel fusion
XX immunoglobulin heavy chain (IGH) protein with a mammalian, especially
XX human, Igh chain fused in frame at its N-terminus to one or more human

CC immune deficiency virus (HIV) gp120 epitopes. Fusion Ig proteins and/or
CC transfected cells are used to tolerate subjects to gp120 epitopes and to
CC maintain this tolerance, particularly for treatment of HIV infection,
CC optionally together with other therapeutic/prophylactic agents such as
CC vaccines, chemotherapeutic agents and immune response modifiers. Such
CC proteins can be used against other diseases where an immune response is
CC deleterious, e.g. microbial infection, tumours or autoimmune disease.
CC Induction of tolerance suppresses production of antibodies against gp120,
CC so prevents or inhibits 'bystander' apoptosis of uninfected T cells that
CC are bound to gp120 protein, maximising induction of protective antiviral
CC T cell immunity.
XX
XX Sequence 3 AA:
SQ

Query Match 25.0%; Score 2; DB 19; Length 3;
Best Local Similarity 100.0%; Pred. No. 4.3e+05; Mismatches 0; Indels 0; Gaps 0;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
II
DB 2 af 3

RESULT 20
AAW84197
ID AAW84197 standard; peptide; 3 AA.
XX
AC AAW84197;
XX
DT 25-MAR-1999 (first entry)
XX
DE Peptide comprising a proteinase site.
XX
KW Proteinase site; bone morphogenetic fusion protein; bone binding site;
KW bone morphogenetic protein; transforming growth factor beta;
KW active fragment; wound healing; bone growth.
XX
OS Unidentified.
XX
XX WO9855137-A1.
XX
PD 10-DEC-1998.
XX
XX 02-JUN-1998; 98WO-US11189.
XX
PR 03-JUN-1997; 97US-0868452.
XX
XX (HALL/) HALL F L.
XX
PA (HANB/) HAN B.
PA (NIMN/) NIMNI M E.
PA (SHOR/) SHORS E C.
PA (WU L/) WU L.
XX
XX Hall FL, Han B, Nimni ME, Shors EC, Wu L;
XX
XX WPI: 1999-059875/05.
XX
DR New bone morphogenetic fusion proteins - comprising a purification
XX tag and a bone morphogenetic active fragment, used for enhancing
XX wound healing or bone growth
XX
XX Claim 8; Page 35; 64pp; English.
XX
XX The present peptide represents a proteinase site used in the creation
XX of the bone morphogenetic fusion proteins of the invention. The bone
XX morphogenetic fusion protein may contain some or all of the following
XX elements: a purification tag, a proteinase site, an ECM/bone binding
XX site, a second proteinase site, and a bone morphogenetic protein
XX active fragment. The fusion proteins of the invention also includes
XX proteins that have transforming growth factor beta active fragments
XX instead of bone morphogenetic protein active fragments. The bone
XX morphogenetic fusion proteins can be used for enhancing wound healing

CC or bone growth.
XX
SO Sequence 3 AA;

Query Match 25.0%; Score 2; DB 20; Length 3;
Best Local Similarity 100.0%; Pred. No. 4.3e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
11
Db 2 af 3

RESULT 21

AAI49387
ID AAY49387 standard; peptide; 3 AA.

AC AAY49387;

XX 13-MAR-2000 (first entry)

DE Dermorphin derivative peptide as diagnostic marker.

KW Diagnostic marker; human disorder; opiate; autism spectral disorder;

KW autism pervasive developmental disorder; Aspergers syndrome; dermorphin;

KW attention deficient disorder; attention hyperactivity disorder;

KW multiple sclerosis; Parkinson's disease; Alzheimer's dementia.

XX Synthetic.

OS Homo sapiens.

XX EP969015-A2.

XX 05-JAN-2000.

XX 15-JUN-1999; 99EP-0304636.

XX 15-JUN-1998; 98US-0089237.

PR 15-JUN-1998; 98US-0089238.

PR 24-MAY-1999; 99US-0317702.

XX (ORTH-) ORTHO-CLINICAL DIAGNOSTICS INC.

XX Shanahan MR, Venturini AJ, Daisis JL, Friedman AE;

XX WPI: 2000-074781/07.

XX Diagnosing human disorders e.g. autism spectral disorders, multiple

XX sclerosis, Parkinson's disease and Alzheimer's dementia -

XX Claim 6; Page 6; 44pp; English.

XX The invention provides diagnostic markers for a human disorder,

XX comprising either opiate-like peptides or opiate-derived peptides. The

XX novel peptides are used as diagnostic marker, in ex-vivo methods of

XX diagnosing human disorders e.g. autism spectral disorders including

XX autism pervasive developmental disorder, Aspergers syndrome, attention

XX deficient disorder and attention hyperactivity disorder, and multiple

XX sclerosis, Parkinson's disease and Alzheimer's dementia. Sequences

XX AA49371-413 represent peptide diagnostic markers for the human

XX disorders specified above.

SO Sequence 3 AA;

OY 5 af 6

11

Db 2 af 3

RESULT 22

AAI0545
ID AAI0545 standard; peptide; 4 AA.

XX AAI0545;

XX 22-DEC-1992 (first entry)

DE Analgesic peptide #1.

XX Growth Hormone release; prolactin; painkiller; anti-psychotic.

XX Synthetic.

XX key Location/Qualifiers

XX Misc-difference 2 /note= "D-Ala"

XX Modified-site 4 /note= "Gly-OH, -NH2, -OMe, -NH-NH2,

XX -NH-NH-2, -NH-NH-CO-(CH2)2-CH3,

XX -NHNH-1auryl, -NHNH-benzoyl,

XX -NHNHAdoc (Ad = adamantyl)

XX or NNNHBOC"

XX BE885283-A.

XX 18-MAR-1981.

XX 19-SEP-1980; 80BE-0983143.

XX 09-MAY-1980; 80GB-0015412.

PR 20-SEP-1979; 79GB-0032580.

PR 20-SEP-1978; 78GB-0032580.

PR 17-SEP-1980; 80GB-0029999.

XX (FARM) FARMITAL ERBA C SPA.

XX Decastiglione R, Faoro F, Perseo G, Piani S, Santangelo F;

XX WPI: 1981-23404D/14 (23404D).

XX Analgesic and antipsychotic polypeptide(s) - also useful in neuro

XX endocrinology and to stimulate liberation of growth hormone and

XX prolactin

XX Example; Page 11; 41pp; French.

XX Peptides covered by this sequence are preferred examples of a highly

XX generic formula for peptides with analgesic and antipsychotic

XX properties. The peptides also have Growth Hormone releasing and

XX prolactin releasing properties. The peptides whose C-termini are

XX protected by -(NH)2-CO-(CH2)2-CH3, -NHNH2, -NHNH1 or -NHNH1 are

XX in the form of their hydrochloride salts and the peptide whose

XX C-terminus is protected by -NH-NH3 is in the form of its

XX dihydrochloride salt. See also AAI0546-P10584.

SO Sequence 4 AA;

Query Match 25.0%; Score 2; DB 2; Length 4;

Best Local Similarity 100.0%; Pred. No. 4.3e+05;

Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6

11

Db 2 af 3

RESULT 23

AAI0584
ID AAI0584 standard; peptide; 4 AA.

```
XX AC AAP10584;
XX FT Modified-site 1 Location/Qualifiers
DT 22-DEC-1992 (first entry) /note= "Boc-Tyr"
XX DE Analgesic peptide #40.
XX KW Growth Hormone release; prolactin; painkiller; anti-psychotic.
XX OS Synthetic.
XX FH Key
FT Modified-site 1 Location/Qualifiers
FT Misc-difference 2 /note= "Boc-Tyr"
FT Modified-site 4 /note= "D-Ala"
FT FT /note= "Gly-NH-NH-2, -NH-NH-CO-(CH2)2-CH3
FT FT or -NH-NH2"
XX PN BE885283-A.
XX PD 18-MAR-1981.
XX PF 19-SEP-1980; 80BE-0983143.
XX PR 09-MAY-1980; 80GB-0015412.
XX PR 20-SEP-1979; 79GB-0032590.
XX PR 20-SEP-1978; 78GB-0032590.
XX PR 17-SEP-1980; 80GB-0029999.
XX PA (FARM ) FARMITAL ERBA C SPA.
XX PI Decastiglione R, Faoro F, Perseo G, Piani S, Santangelo F;
XX DR WPI: 1981-23404D/14 (23404D).
XX PT Analgesic and antipsychotic polypeptide(s) - also useful in neuro
XX PT endocrinology and to stimulate liberation of growth hormone and
XX PT prolactin
XX PS Example; Page 11; 41pp; French.
XX CC Peptides covered by this sequence are preferred examples of a highly
XX CC generic formula for peptides with analgesic and antipsychotic
XX CC properties. The peptides also have Growth Hormone releasing and
XX CC prolactin releasing properties. See also AAP10545-P10583.
XX SQ Sequence 4 AA;

Query Match 25.0%; Score 2; DB 2; Length 4;
Best Local Similarity 100.0%; Pred. No. 4.3e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 af 6
II
DB 2 af 3

RESULT 24
AAP10547
ID AAP10547 standard; peptide; 4 AA.
XX
XX AAP10547;
XX
XX 22-DEC-1992 (first entry)
XX
XX Analgesic peptide #3.
XX
XX Growth Hormone release; prolactin; painkiller; anti-psychotic.
XX
XX Synthetic.
```

```
XX FH Key
XX FT Modified-site 1 Location/Qualifiers
XX FT Misc-difference 2 /note= "opt. Boc-protected"
XX FT FT /note= "D-Ala"
XX FT Modified-site 4 /note= "Gly-NH-NH-2 or -NH-NH2"
XX PN BE885283-A.
XX PD 18-MAR-1981.
XX PF 19-SEP-1980; 80BE-0983143.
XX PR 09-MAY-1980; 80GB-0015412.
XX PR 20-SEP-1979; 79GB-0032590.
XX PR 20-SEP-1978; 78GB-0032590.
XX PR 17-SEP-1980; 80GB-0029999.
XX PA (FARM ) FARMITAL ERBA C SPA.
XX PI Decastiglione R, Faoro F, Perseo G, Piani S, Santangelo F;
XX DR WPI: 1981-23404D/14 (23404D).
XX PT Analgesic and antipsychotic polypeptide(s) - also useful in neuro
XX PT endocrinology and to stimulate liberation of growth hormone and
XX PT prolactin
XX PS Example; Page 11; 41pp; French.
XX CC Peptides covered by this sequence are preferred examples of a highly
XX CC generic formula for peptides with analgesic and antipsychotic
XX CC properties. The peptides also have Growth Hormone releasing and
XX CC prolactin releasing properties. See AAP10545-P10584.
XX SQ Sequence 4 AA;

Query Match 25.0%; Score 2; DB 2; Length 4;
Best Local Similarity 100.0%; Pred. No. 4.3e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 af 6
II
DB 2 af 3

RESULT 25
AAP10548
ID AAP10548 standard; peptide; 4 AA.
XX
XX AAP10548;
XX
XX 22-DEC-1992 (first entry)
XX
XX Analgesic peptide #4.
XX
XX Growth Hormone release; prolactin; painkiller; anti-psychotic.
XX
XX Synthetic.
XX FH Key
XX FT Modified-site 1 Location/Qualifiers
XX FT Misc-difference 2 /note= "opt. Boc-protected"
XX FT Modified-site 3 /note= "D-Ala"
XX FT Modified-site 4 /label= MeGly
XX FT Modified-site 4 /note= "Gly-NH-NH2 or Gly-NHNH2"
XX
```

PN BE885283-A.
XX
PD 18-MAR-1981.
XX
PF 19-SEP-1980; 80BE-0983143.
XX
PR 09-MAY-1980; 80GB-0015412.
PR 20-SEP-1979; 79GB-0032590.
PR 20-SEP-1978; 78GB-0032590.
PR 17-SEP-1980; 80GB-0029999.
XX
PA (FARM) FARMITAL ERBA C SPA.
XX
PI Decastiglione R, Faoro F, Perseo G, Piani S, Santangelo F;
XX WPI: 1981-23404D/14 (23404D).
XX
XX Analgesic and antipsychotic polypeptide(s) - also useful in neuro
PT endocrinology and to stimulate liberation of growth hormone and
PT prolactin
XX
XX Example; Page 11; 41pp; French.
PS
CC Peptides covered by this sequence are preferred examples of a highly
CC generic formula for peptides with analgesic and antipsychotic
CC properties. The peptides also have Growth Hormone releasing and
CC prolactin releasing properties. When the N-terminus is not
CC Boc-protected, the peptides are in the form of their hydrochloride
CC (i.e. -NHNH2) or dihydrochloride (i.e. -NHNH2) salts.
CC See AAP10545-P10584.
XX
SQ Sequence 4 AA;

Query Match 25.0%; Score 2; DB 2; Length 4;
Best Local Similarity 100.0%; Pred. No. 4.3e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
II
Db 2 af 3

Search completed: January 14, 2002, 07:56:31
Job time: 726 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: January 14, 2002, 07:57:31 ; Search time 41.59 Seconds
(without alignments)
4.329 Million cell updates/sec

Title: 09-185908-1D

Perfect score: 8

Sequence: 1 wkxxafxg 8

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
212252 seqs, 22503292 residues

Searched:

Word size : 0

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/prodata/2/1aa/backfile1.pep.*

Prdcd. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3	37.5	15	2	US-08-935-886-1	Sequence 1, Appl
2	3	37.5	24	2	US-08-404-531B-34	Sequence 34, Appl
3	3	37.5	24	3	US-08-476-900A-34	Sequence 34, Appl
4	3	37.5	24	3	US-08-488-546A-34	Sequence 34, Appl
5	3	37.5	112	4	US-08-905-223-302	Sequence 302, Appl
6	2	25.0	3	1	US-08-213-897A-13	Sequence 13, Appl
7	2	25.0	3	1	US-08-213-897A-14	Sequence 14, Appl
8	2	25.0	3	1	US-08-470-837-11	Sequence 11, Appl
9	2	25.0	3	4	US-09-461-697-405	Sequence 405, Appl
10	2	25.0	4	1	US-07-719-692A-1	Sequence 1, Appl
11	2	25.0	4	1	US-07-719-692A-2	Sequence 2, Appl
12	2	25.0	4	1	US-07-719-692A-3	Sequence 3, Appl
13	2	25.0	4	1	US-07-805-727-8	Sequence 8, Appl
14	2	25.0	4	1	US-08-079-445-3	Sequence 3, Appl
15	2	25.0	4	1	US-07-840-077A-7	Sequence 7, Appl
16	2	25.0	4	1	US-07-866-018-4	Sequence 4, Appl
17	2	25.0	4	1	US-08-127-904-8	Sequence 8, Appl
18	2	25.0	4	1	US-08-390-272-8	Sequence 8, Appl
19	2	25.0	4	1	US-08-102-372-4	Sequence 4, Appl
20	2	25.0	4	1	US-08-206-789-5	Sequence 5, Appl
21	2	25.0	4	1	US-08-211-070A-7	Sequence 7, Appl
22	2	25.0	4	1	US-08-176-938-24	Sequence 24, Appl
23	2	25.0	4	1	US-08-176-938-35	Sequence 35, Appl
24	2	25.0	4	1	US-08-193-977-20	Sequence 20, Appl
25	2	25.0	4	1	US-08-461-611-12	Sequence 12, Appl
26	2	25.0	4	1	US-08-340-045-18	Sequence 18, Appl
27	2	25.0	4	1	US-08-340-045-20	Sequence 20, Appl

28	2	25.0	4	1	US-08-454-950-7	Sequence 7, Appl
29	2	25.0	4	1	US-08-224-868-7	Sequence 7, Appl
30	2	25.0	4	1	US-08-434-761-3	Sequence 3, Appl
31	2	25.0	4	1	US-08-338-890B-1	Sequence 1, Appl
32	2	25.0	4	1	US-08-561-478-4	Sequence 4, Appl
33	2	25.0	4	1	US-08-549-008-29	Sequence 29, Appl
34	2	25.0	4	1	US-08-387-156-23	Sequence 23, Appl
35	2	25.0	4	1	US-08-425-069-18	Sequence 18, Appl
36	2	25.0	4	1	US-08-460-343B-63	Sequence 63, Appl
37	2	25.0	4	1	US-08-388-321-8	Sequence 8, Appl
38	2	25.0	4	1	US-08-466-632-8	Sequence 8, Appl
39	2	25.0	4	1	US-08-446-177-8	Sequence 8, Appl
40	2	25.0	4	1	US-08-454-949-7	Sequence 7, Appl
41	2	25.0	4	1	US-08-643-709A-4	Sequence 4, Appl
42	2	25.0	4	1	US-08-398-028B-63	Sequence 63, Appl
43	2	25.0	4	1	US-08-798-897-46	Sequence 46, Appl
44	2	25.0	4	1	US-08-877-460-4	Sequence 4, Appl
45	2	25.0	4	1	US-08-456-424-22	Sequence 22, Appl
46	2	25.0	4	1	US-08-456-424-23	Sequence 23, Appl
47	2	25.0	4	1	US-08-456-424-24	Sequence 24, Appl
48	2	25.0	4	1	US-08-456-424-93	Sequence 93, Appl
49	2	25.0	4	1	US-08-456-424-94	Sequence 94, Appl
50	2	25.0	4	1	US-08-456-424-95	Sequence 95, Appl

ALIGNMENTS

RESULT 1
US-08-935-886-1
Sequence 1, Application US/08935886
Patent No. 5994625
GENERAL INFORMATION:
APPLICANT: Melchers, Leo S.
APPLICANT: Sela-Buurlage, Marianne B.
APPLICANT: Bres-Vloemans, Alexandra A.
APPLICANT: Ponsteln, Anne S.
APPLICANT: Apotheker-de Groot, Marlon
APPLICANT: Cornelissen, Bernardus J. C.
TITLE OF INVENTION: Antifungal Cytlin Binding Proteins and DNA
TITLE OF INVENTION: Coding Therefor
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ladas & Parry
STREET: 26 West 61st Street
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10023
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 14.4 MB storage
COMPUTER: IBM PC/XT/AT or compatibles
OPERATING SYSTEM: DOS 6.20
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/935,886
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/411,640
FILING DATE: 05-APR-1995
APPLICATION NUMBER: PCT/EP93/02790
FILING DATE: 05-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mass, Clifford J.
REGISTRATION NUMBER: 30,086
REFERENCE/DOCKET NUMBER: 0-010139-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 708-1800
TELEFAX: (212) 246-8959
TELEX: 233288
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

```

: LENGTH: 15 Amino Acids
: TYPE: Amino Acid
: TOPOLOGY: Linear
: MOLECULE TYPE: Protein
: HYPOTHETICAL: NO
: FRAGMENT TYPE: Internal
: ORIGINAL SOURCE:
: ORGANISM: Nicotiana tabacum
: STRAIN: Samsun NN
: DEVELOPMENTAL STAGE: Mature
: TISSUE TYPE: Leaf, wounded
US-08-935-886-1

Query Match          37.5%; Score 3; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 afxg 8
    ||||
Db 3 AFXG 6

RESULT 2
US-08-404-531B-34
: Sequence 34, Application US/08404531B
: Patent No. 5863724
: GENERAL INFORMATION:
: APPLICANT: Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson, Pamela
: APPLICANT: Thomas, Gilbert Cole, and Robert Gagel
: TITLE OF INVENTION: Sequence Encoding Mammalian Sulfonylurea Receptor
: Patent No. 5863724
: NUMBER OF SEQUENCES: 49
: CORRESPONDENCE ADDRESS:
: ADDRESSSEE: Woodcock Washburn Kurtz Mackiewicz &
: ADDRESSSEE: No. 5863724is
: STREET: One Liberty Place 46th. Floor
: CITY: Philadelphia
: STATE: PA
: COUNTRY: USA
: ZIP: 19103
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/404,531B
: FILING DATE: 15-MAR-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Beardell, Lori Y.
: REGISTRATION NUMBER: 34,293
: REFERENCE/DOCKET NUMBER: BYLR-0003
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 215-568-3100
: TELEFAX: 215-568-3439
: INFORMATION FOR SEQ ID NO: 34:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 24 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
US-08-404-531B-34

Query Match          37.5%; Score 3; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 afxg 8
    ||||
Db 3 AFXG 6
```

```

Db 3 AFXG 6

RESULT 3
US-08-476-900A-34
: Sequence 34, Application US/08476900A
: Patent No. 6031150
: GENERAL INFORMATION:
: APPLICANT: Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson
: APPLICANT: Thomas, Gilbert Cole, and Robert Gagel
: TITLE OF INVENTION: Sequence Encoding Mammalian Sulfonylurea Receptor
: Patent No. 6031150
: TITLE OF INVENTION: and Method of Detecting Persistent Hyperinsulinemic Hypogly
: TITLE OF INVENTION: Infancy
: NUMBER OF SEQUENCES: 49
: CORRESPONDENCE ADDRESS:
: ADDRESSSEE: Woodcock Washburn Kurtz Mackiewicz &
: ADDRESSSEE: No. 6031150is
: STREET: One Liberty Place 46th. Floor
: CITY: Philadelphia
: STATE: PA
: COUNTRY: USA
: ZIP: 19103
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/476,900A
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 800
: ATTORNEY/AGENT INFORMATION:
: NAME: Beardell, Lori Y.
: REGISTRATION NUMBER: 34,293
: REFERENCE/DOCKET NUMBER: BYLR-0027
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 215-568-3100
: TELEFAX: 215-568-3439
: INFORMATION FOR SEQ ID NO: 34:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 24 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
US-08-476-900A-34

Query Match          37.5%; Score 3; DB 3; Length 24;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 afxg 8
    ||||
Db 3 AFXG 6

RESULT 4
US-08-488-546A-34
: Sequence 34, Application US/08488546A
: Patent No. 6054313
: GENERAL INFORMATION:
: APPLICANT: Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson, Pamela
: APPLICANT: Thomas, Gilbert Cole, and Robert Gagel
: TITLE OF INVENTION: Sequence Encoding Mammalian Sulfonylurea Receptor
: Patent No. 6054313
: NUMBER OF SEQUENCES: 49
: CORRESPONDENCE ADDRESS:
: ADDRESSSEE: Woodcock Washburn Kurtz Mackiewicz &
: ADDRESSSEE: No. 6054313is
: STREET: One Liberty Place 46th. Floor
```

CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,546A
FILING DATE: 07-JUNE-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/404,531
FILING DATE: 15-MARCH-1995
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Beardell, Lori Y.
REGISTRATION NUMBER: 34,293
REFERENCE/DOCKET NUMBER: BYLR-0026
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 24 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-488-546A-34

Query Match
Best Local Similarity 37.5%; Score 3; DB 3; Length 24;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 afxg 8
Db 3 AFXG 6

RESULT 5
US-08-905-223-302
Sequence 302, Application US/08905223
Patent No. 6222029
GENERAL INFORMATION:
APPLICANT: Edwards, Jean-Baptiste D.
APPLICANT: Duclert, Aymeric
APPLICANT: Lacroix, Bruno
TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS
NUMBER OF SEQUENCES: 503
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knodbe, Martens, Olson & Bear
STREET: 501 West Broadway
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101-3505
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Win95
SOFTWARE: Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/905,223
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Nod A.
REGISTRATION NUMBER: 29,655

REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 302:
SEQUENCE CHARACTERISTICS:
LENGTH: 112 amino acids
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
ORIGINAL SOURCE:
ORGANISM: Homo Sapiens
TISSUE TYPE: Brain
FEATURE:
NAME/KEY: sig-peptide
LOCATION: -52...-1
IDENTIFICATION METHOD: Von Heljne matrix
OTHER INFORMATION: score 7
US-08-905-223-302

Query Match
Best Local Similarity 37.5%; Score 3; DB 4; Length 112;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 afxg 8
Db 64 AFXG 67

RESULT 6
US-08-213-897A-13
Sequence 13, Application US/08213897A
Patent No. 5618790
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Protease Mediated Drug Delivery System
NUMBER OF SEQUENCES: 18
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/213,897A
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/593,867
FILING DATE: 05-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/833,183
FILING DATE: 10-FEB-1992
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 3 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-213-897A-13

Query Match
Best Local Similarity 25.0%; Score 2; DB 1; Length 3;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 af 6
Db 1 Af 2

RESULT 7
US-08-213-897A-14

; Sequence 14, Application US/08213897A
; Patent No. 5618790
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Protease Mediated Drug Delivery System
; NUMBER OF SEQUENCES: 18
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/213,897A
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/593,867
; FILING DATE: 05-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/833,183
; FILING DATE: 10-FEB-1992
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-213-897A-14

Query Match 25.0%; Score 2; DB 1; Length 3;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
11
1 AF 2

RESULT 8
US-08-470-837-11
; Sequence 11, Application US/08470837
; Patent No. 5800811
; GENERAL INFORMATION:
; APPLICANT: Nimni, Marcel E.
; APPLICANT: Hall, Frederick L.
; APPLICANT: Tuan, Tai-Lan
; APPLICANT: Wu, Linglao
; APPLICANT: Cheung, David T.
; TITLE OF INVENTION: Transforming Growth Factor B Fusion
; TITLE OF INVENTION: and
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 1150 Santa Monica Boulevard, Suite 400
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90025-3395
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,837
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharp, Janice A.
; REGISTRATION NUMBER: 34,051
; REFERENCE/DOCKET NUMBER: 30630-1US01
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 310-445-1140
; TELEFAX: 310-445-9031
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-470-837-11

Query Match 25.0%; Score 2; DB 1; Length 3;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
11
2 AF 3

RESULT 9
US-09-461-697-405
; Sequence 405, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Purnam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 405
; LENGTH: 3
; TYPE: PPT
; ORGANISM: Homo sapiens
US-09-461-697-405

Query Match 25.0%; Score 2; DB 4; Length 3;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
11
2 WK 3

RESULT 10
US-07-719-692A-1
; Sequence 1, Application US/07719692A
; Patent No. 5250414
; GENERAL INFORMATION:
; APPLICANT: Schwab, Martin E. et al.
; TITLE OF INVENTION: Neutite Growth Regulatory Factors
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/719,692A
FILING DATE: 19910624
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: MISTOCK, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 6526-022
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 8698864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /note= "amide at carboxy terminus"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /note= "carboxy derivatized"
OTHER INFORMATION: N-terminus"
US-07-719-692A-1

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
||
Db 2 Af 3

RESULT 11
US-07-719-692A-2
Sequence 2, Application US/07719692A
Patent No. 5250414
GENERAL INFORMATION:
APPLICANT: Schwab, Martin E. et al.
TITLE OF INVENTION: Neurite Growth Regulatory Factors
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/719,692A
FILING DATE: 19910624
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: MISTOCK, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 6526-022
TELECOMMUNICATION INFORMATION:

TELEPHONE: 212 790-9090
TELEFAX: 212 8698864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /note= "carboxy derivatized"
OTHER INFORMATION: end"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /note= "tyrosine is radiolabeled"
OTHER INFORMATION: with Iodine-125."
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /note= "amide at carboxy-terminus"
US-07-719-692A-2

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
||
Db 2 Af 3

RESULT 12
US-07-719-692A-3
Sequence 3, Application US/07719692A
Patent No. 5250414
GENERAL INFORMATION:
APPLICANT: Schwab, Martin E. et al.
TITLE OF INVENTION: Neurite Growth Regulatory Factors
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/719,692A
FILING DATE: 19910624
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: MISTOCK, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 6526-022
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 8698864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single

TOPOLOGY: unknown
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /note= "amide at carboxy-terminus"
US-07-719-692A-3

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
11
Db 2 Af 3

RESULT 13
US-07-805-727-8
Sequence 8, Application US/07805727
Patent No. 5424186
GENERAL INFORMATION:
APPLICANT: Fodor, Stephen P.A.
APPLICANT: Stryer, Lubert
APPLICANT: Pirrung, Michael C.
APPLICANT: Read, J. Leighton
TITLE OF INVENTION: Very large Scale Immobilized Polymer
TITLE OF INVENTION: Synthesis
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESS: Vernon A. No. 5424186v1el
STREET: One Market Plaza, Steuart Tower, Suite 2000
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/805,727
FILING DATE: 19911206
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: No. 5424186v1el, Vernon A.
REGISTRATION NUMBER: 32,483
REFERENCE/DOCKET NUMBER: 11509A1)1)1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-07-805-727-8

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
11
Db 3 Af 4

RESULT 14.

US-08-079-445-3
Sequence 3, Application US/08079445
Patent No. 5440016
GENERAL INFORMATION:
APPLICANT: Blondelle, Sylvie E.
APPLICANT: Pinilla, Clemencia
APPLICANT: Eichler, Julia
APPLICANT: Houghten, Richard A.
TITLE OF INVENTION: PEPTIDES HAVING ANTI-MICROBIAL,
TITLE OF INVENTION: HEMOLYTIC, ANTI-MELITTIN OR ANTI-ENZYMATIC ACTIVITY
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: CAMPBELL AND FLORES
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States of America
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/079,445
FILING DATE: 18-JUN-1993
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-HP 9648
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
TOPOLOGY: linear
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /label= Xaa
OTHER INFORMATION: /note= "Xaa=Kfmoc"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /label= Xaa
OTHER INFORMATION: /note= "Xaa=any amino acid"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /note= "C-terminal amino acid is
OTHER INFORMATION: amidated"
US-08-079-445-3

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
11
Db 2 wk 3

RESULT 15
US-07-840-077A-7
Sequence 7, Application US/07840077A
Patent No. 5443816
GENERAL INFORMATION:
APPLICANT: Zamora, Paul O.

APPLICANT: Rhodes, Buck A.
TITLE OF INVENTION: Peptide-Metal Ion
NUMBER OF SEQUENCES: 7
CURRENT APPLICATION DATA:
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rhomed Incorporated
STREET: 4261 Balloon Park
CITY: Albuquerque
STATE: NM
COUNTRY: U.S.A.
ZIP: 87109-5802
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 Inch, 1.44 MB
COMPUTER: IBM PC/XT/AT, IBM PS/2 or
COMPILER: compatibles
OPERATING SYSTEM: PC-DOS or MS-DOS
SOFTWARE: Wordperfect 6.0a for Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/840,077A
FILING DATE: 20-FEB-1992
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/816,477
FILING DATE: 03-JAN-1992
ATTORNEY/AGENT INFORMATION:
NAME: Deborah A. Peacock
REGISTRATION NUMBER: 31,649
TELECOMMUNICATION INFORMATION:
TELEPHONE: (505) 242-9677
TELEFAX: (505) 243-2542
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-07-840-077A-7

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wk 2
Db 2 wk 3

RESULT 16
US-07-866-018-4
Sequence 4, Application US/07866018
Patent No. 5470705
GENERAL INFORMATION:
APPLICANT: Grossman, Paul D.
APPLICANT: Fung, Steven
APPLICANT: Menchen, Steven M.
APPLICANT: Woo, Sam L.
APPLICANT: Winn-Deen, Emily S.
TITLE OF INVENTION: Probe Composition and Method
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Peter Dehlinger
STREET: 350 Cambridge Avenue, Suite 300
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/866,018
FILING DATE: 19920407
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 0550-0023.30
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0960
TELEFAX: (415) 324-0860
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: TETRAPEPTIDE, PAGE 28
US-07-866-018-4

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 af 6
Db 2 AF 3

RESULT 17
US-08-127-904-8
Sequence 8, Application US/08127904
Patent No. 5470951
GENERAL INFORMATION:
APPLICANT: Eugene Roberts
TITLE OF INVENTION: Method For Antagonizing
TITLE OF INVENTION: Amnesic Effects of Amyloid n
TITLE OF INVENTION: Protein and Improving the
TITLE OF INVENTION: Quality of Life in Individuals
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: City of Hope
STREET: 1500 East Duarte Road
CITY: Duarte
STATE: California
COUNTRY: United States of America
ZIP: 91010-0269
COMPUTER READABLE FORM:
MEDIUM TYPE: 3M Double Density 5 1/4" diskette
COMPUTER: Wang PC
OPERATING SYSTEM: MS DOS Version 3.20
SOFTWARE: Microsoft
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/127,904
FILING DATE: 29 September 1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA: No. 5470951e
ATTORNEY/AGENT INFORMATION:
NAME: Itons, Edward S.
REGISTRATION NUMBER: 16,541
REFERENCE/DOCKET NUMBER: No. 5470951e
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 783-6040
TELEFAX: (202) 783-6031
TELEX: No. 5470951e
INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 4
TYPE: Amino Acid
STRANDEDNESS:
TOPOLOGY: Unknown
US-08-127-904-8

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;

Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
11
Db 1 AF 2

RESULT 18

US-08-390-272-8
Sequence 8, Application US/08390272
Patent No. 5489678

GENERAL INFORMATION:

APPLICANT: Fodor, Stephen P.A.
APPLICANT: Strayer, Lubert
APPLICANT: Winkler, James L.
APPLICANT: Holmes, Christopher P.
APPLICANT: Solas, Dennis W.
TITLE OF INVENTION: Very Large Scale Immobilized Polymer
TITLE OF INVENTION: Synthesis
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Vernon A. No. 5489678v1e1
STREET: One Market Plaza, Steuart Tower, Suite 2000
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94105

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/390,272
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/624,120
FILING DATE: 06-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: No. 5489678v1e1, Vernon A.
REGISTRATION NUMBER: 32,483
REFERENCE/DOCKET NUMBER: 11509-28
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-390-272-8

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
11
Db 3 AF 4

RESULT 19

US-08-102-372-4
Sequence 4, Application US/08102372
Patent No. 5514543

GENERAL INFORMATION:

APPLICANT: Grossman, Paul D.
APPLICANT: Pung, Steven
APPLICANT: Menchen, Steven M.
APPLICANT: Woo, Sam L.
APPLICANT: Winn-Deen, Emily S.
TITLE OF INVENTION: Probe Composition and Method
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Peter Dehlinger
STREET: 350 Cambridge Avenue, Suite 300
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/102,372
FILING DATE: 04-AUG-1993
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Powers, Vincent M.
REGISTRATION NUMBER: 36,246
REFERENCE/DOCKET NUMBER: 0550-0023.31
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: synthetic tetrapeptide
US-08-102-372-4

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
11
Db 2 AF 3

RESULT 20

US-08-206-789-5
Sequence 5, Application US/08206789
Patent No. 5580854

GENERAL INFORMATION:

APPLICANT: Orłowski, Marian
APPLICANT: Cardozo, Christopher
APPLICANT: Vinitsky, Alexander
TITLE OF INVENTION: SUBSTRATE-RELATED
TITLE OF INVENTION: PEPTIDYL-ALDEHYDE
TITLE OF INVENTION: INHIBITORS OF THE PROTEOLYTIC ACTIVITY OF THE
MULTICATALYTIC PROTEINASE COMPLEX
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:

ADDRESSEE: BRUMBAUGH, GRAVES, DONOHUE &
ADDRESSER: RAYMOND
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112-2500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/206,789
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Seide, Rochelle K.
REGISTRATION NUMBER: 32,300
REFERENCE/DOCKET NUMBER: A29525 - 165/25989
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-408-2626
TELEFAX: 212-765-2519
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..4
US-08-206-789-5

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05; Indels 0;
Matches 2; Conservative 0; Mismatches 0; Gaps 0;

Qy 5 af 6
11
Db 3 af 4

RESULT 21
US-08-211-070A-7
Sequence 7, Application US/08211070A
Patent No. 5585350
GENERAL INFORMATION:
APPLICANT: Friedlich, Thomas
APPLICANT: Koewer, Wolfgang
APPLICANT: Kroeger, Burkhard
APPLICANT: Bialojan, Siegfried
TITLE OF INVENTION: No. 5585350e1 thrombin-inhibitory protein from ticks.
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSER: Kell & Weinlauf
STREET: 1101 Connecticut Avenue
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 KB storage
COMPUTER: IBM AT-compatible, 80486 processor
OPERATING SYSTEM: MS-DOS version 6.0
SOFTWARE: Wordperfect version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/211,070A
FILING DATE: 18-MAR-1994
CLASSIFICATION: 435
CLASSIFICATION: C12P 21/02
CLASSIFICATION: A61K 37/64

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP92/02198
FILING DATE: 23-SEP-1992
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-211-070A-7

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05; Indels 0;
Matches 2; Conservative 0; Mismatches 0; Gaps 0;

Qy 5 af 6
11
Db 1 af 2

RESULT 22
US-08-176-938-24
Sequence 24, Application US/08176938
Patent No. 5602099
GENERAL INFORMATION:
APPLICANT: Schiller, Peter W.
TITLE OF INVENTION: New Peptides
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: White & Case
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: U.S.A.
ZIP: 10036-2787
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/176,938
FILING DATE: 04-JAN-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Sterner Ph.D., Richard J.
REGISTRATION NUMBER: 35,372
REFERENCE/DOCKET NUMBER: 1103326-080
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-819-8783
TELEFAX: 212-354-8113
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 2
OTHER INFORMATION: /label= Tic
OTHER INFORMATION: /note= "1,2,3,4-tetrahydroisouinolnline-3-carboxylic acid"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 3
OTHER INFORMATION: /label= 2-Na1
OTHER INFORMATION: /note= "3-(2'-naphthyl)alanine"
US-08-176-938-24

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05; Indels 0;
Matches 2; Conservative 0; Mismatches 0; Gaps 0;

OY 5 af 6
11
Db 3 AF 4

RESULT 23
US-08-176-938-35
; Sequence 35, Application US/08176938
; Patent No. 5602099
; GENERAL INFORMATION:
; APPLICANT: Schiller, Peter W.
; TITLE OF INVENTION: New Peptides
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/176,938
; FILING DATE: 04-JAN-1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Steiner Ph.D., Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-080
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-819-8783
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 2
; OTHER INFORMATION: /label= TIC
; OTHER INFORMATION: /note= "1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 3
; OTHER INFORMATION: /label= 1-Nal
; OTHER INFORMATION: /note= "3-(1'-naphthyl)alanine"
US-08-176-938-35

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
11
Db 3 AF 4

RESULT 24
US-08-193-977-20
; Sequence 20, Application US/08193977
; Patent No. 5625031
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, KEVIN R.
; APPLICANT: COLEMAN, KEVIN G.
; TITLE OF INVENTION: PEPTIDE INHIBITORS OF THE P33CDK2 AND

; TITLE OF INVENTION: P34CDK2 CELL CYCLE REGULATORY KINASES AND HUMAN
; TITLE OF INVENTION: PAPILLOMAVIRUS E7 ONCOPROTEIN
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: REED & ROBINS
; STREET: 635 BRYANT STREET
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/193,977
; FILING DATE: 08-FEB-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: ROBINS, ROBERTA L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 5998-0016
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-193-977-20

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 af 6
11
Db 2 AF 3

RESULT 25
US-08-461-611-12
; Sequence 12, Application US/08461611
; Patent No. 5635477
; GENERAL INFORMATION:
; APPLICANT: William Frank Degrado, Sharon Anne Jackson, Shaker Ahmed Mousa, Anju
; APPLICANT: Parthasarathy, Michael Sworin, Maria Rafalski
; TITLE OF INVENTION: Cyclic Compounds Useful as
; TITLE OF INVENTION: Inhibitors of Platelet Glycoprotein IIb/IIIa
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Dupont Merck Pharmaceutical
; STREET: 1007 Market Street, Legal Department
; CITY: Wilmington
; STATE: DE
; COUNTRY: U.S.
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.50 inch disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple Macintosh
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,611
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA: none

APPLICATION NUMBER: 08/038,448
 FILING DATE: March 29, 1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Ferguson, Blat, O.
 REGISTRATION NUMBER: 34,329
 REFERENCE/DOCKET NUMBER: BP-6543-B
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 302-892-1676
 TELEFAX: 302-892-8536
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4 amino acids
 TYPE: amino acid
 TOPOLOGY: circular
 MOLECULE TYPE: peptide
 ORIGINAL SOURCE: synthetic
 FEATURE:
 OTHER INFORMATION: Example Number 12a;
 OTHER INFORMATION: GP1b/11a Inhibitor
 US-08-461-611-12

Query Match 25.0%; Score 2; DB 1; Length 4;
 Best Local Similarity 100.0%; Pred No. 1.6e+05;
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 fxg 8
 |||
 Db 1 FXG 3

Search completed: January 14, 2002, 07:57:31
 Job time: 376 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:54:50 ; Search time 81.98 Seconds

(without alignments)
14.274 Million cell updates/sec

Title: 09-185908-1A

Perfect score: 8

Sequence: 1 wxxxxyxg 8

Scoring table: OLIGO

Gapop 60.0 , Capext 60.0

Searched: 473505 seqs, 146272329 residues

Word size: 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 50 summaries

Database:

- 1: SPTRMBL.17.*
- 2: SP_archaea.*
- 3: SP_bacteria.*
- 4: SP_fungi.*
- 5: SP_invertebrate.*
- 6: SP_mammal.*
- 7: SP_mhc.*
- 8: SP_organelle.*
- 9: SP_phage.*
- 10: SP_plant.*
- 11: SP_protent.*
- 12: SP_virus.*
- 13: SP_vertebrate.*
- 14: SP_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3	37.5	38	3	09UR79	09ur79 trichoderma
2	3	37.5	77	5	09GVM9	09gvm9 leishmania
3	3	37.5	91	6	09WZE1	09mze1 bos taurus
4	3	37.5	131	13	091947	091947 oryza lat
5	3	37.5	136	5	09GMB2	09gmb2 leishmania
6	3	37.5	141	6	027953	027953 balanopter
7	3	37.5	226	4	09WZ74	09wz74 homo sapien
8	3	37.5	485	5	09GXY3	09gyx3 trypanosoma
9	3	37.5	707	8	09BBS2	09bb52 schenocoph
10	2	25.0	7	8	09S945	09s945 saccharomyc
11	2	25.0	8	4	015898	015898 homo sapien
12	2	25.0	8	8	034909	034909 locusta mig
13	2	25.0	8	10	09S8Z4	09s8z4 spinacia ol
14	2	25.0	8	11	06Z721	06z721 rattus norv
15	2	25.0	9	4	09UCW0	09ucw0 homo sapien
16	2	25.0	9	4	09UMAO	09uma0 homo sapien
17	2	25.0	9	4	09BYF9	09byf9 homo sapien
18	2	25.0	9	6	028093	028093 bos taurus
19	2	25.0	9	11	09QW70	09qwt0 mus musculu

20	2	25.0	9	13	09TA14	09ta14 gallus gall
21	2	25.0	10	8	09XMB4	09xmb4 aegilops ta
22	2	25.0	11	4	09UC46	09uc46 homo sapien
23	2	25.0	11	7	077876	077876 oreochromis
24	2	25.0	11	7	077883	077883 oreochromis
25	2	25.0	11	7	077886	077886 oreochromis
26	2	25.0	11	7	077912	077912 oreochromis
27	2	25.0	11	8	032704	032704 nicotiana t
28	2	25.0	11	8	09G607	09g607 apicalotis
29	2	25.0	11	8	09G359	09g359 japalura fl
30	2	25.0	11	11	09JUE6	09jue6 rattus norv
31	2	25.0	12	4	09UC05	09uc05 homo sapien
32	2	25.0	12	12	083139	083139 barley scrt
33	2	25.0	13	4	09UER3	09uee3 homo sapien
34	2	25.0	13	4	09NR93	09nr93 homo sapien
35	2	25.0	13	8	09XLI2	09xli2 bemisia tab
36	2	25.0	13	8	09THR8	09thr8 bryopsis sp
37	2	25.0	13	12	064813	064813 autographa
38	2	25.0	14	2	052220	052220 salmonella
39	2	25.0	14	2	09R506	09r506 burkholderi
40	2	25.0	14	8	09B696	09b696 hordeum vul
41	2	25.0	14	8	09MT61	09mt61 allium cepa
42	2	25.0	14	8	09MRV4	09mrv4 allium porr
43	2	25.0	14	8	09MRV1	09mrv1 allium satl
44	2	25.0	14	8	09MR78	09mr78 aloce vera.
45	2	25.0	14	8	09MR76	09mr76 hordeum mur
46	2	25.0	14	10	P82433	P82433 nicotiana t
47	2	25.0	14	10	09EPT0	09ept0 allium cepa
48	2	25.0	14	11	070599	070599 rattus norv
49	2	25.0	14	11	09Z0G5	09z0g5 mus musculu
50	2	25.0	14	12	067112	067112 influenza a

ALIGNMENTS

RESULT 1
ID 09UR79 PRELIMINARY; PRT; 38 AA.
AC 09UR79;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE CELLOBIOHYDROLASE II CORE PROTEIN, CBH II Cp-3.2.1.91.
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=51453;
RN [1]
RP SEQUENCE.
RX MEDLINE=94250349; PubMed=8192865;
RA Woodward J., Brown J.P., Evans B.R., Affholter K.A.;
RT "Papain digestion of crude Trichoderma reesei cellulase: purification
RL and properties of cellobiohydrolase I and II core proteins.";
DR Biotechnol. Appl. Biochem. 19:141-153(1994).
HSSP: P07987; ICB2.
SQ SEQUENCE 38 AA; 3896 MW; D3E4CBFBFA834E17 CRC64;

Query Match 37.5%; Score 3; DB 3; Length 38;
Best Local Similarity 100.0%; Pred. No. 5.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	5	syxg	8	
DB	4	syxg	7	
RESULT	2			
Q9GVM9		PRELIMINARY;	PRT;	77 AA.
ID	Q9GVM9			
AC	Q9GVM9;			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			

```

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HYPOTHEITICAL 8.1 KDA PROTEIN (FRAGMENT).
GN LM12.1312.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,
RL Oliver K.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL390114; CAC02854.1; -.
KW Hypothetical protein.
FT NON_TER 1
FT SEQUENCE 77 AA; 8080 MW; CDBCf973C60CA844 CRC64;

Query Match
Best Local Similarity 37.5%; Score 3; DB 5; Length 77;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 syxg 8
Db 46 SYXG 49

RESULT 3
O9MZE1 PRELIMINARY; PRT; 91 AA.
ID O9MZE1
AC O9MZE1
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GLUTATHIONE PEROXIDASE (FRAGMENT).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Vermette L., Siroids J., Carriere P., Price C., Silversides D.W.,
RL Lussier J.G.;
RT *Cow glutathione peroxidase.*;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF236854; AAF74263.1; -.
DR InterPro: IPR000889; Glut_peroxidase.
DR Pfam: PF00255; GSHPx; 1.
KW Peroxidase.
FT NON_TER 91
FT SEQUENCE 91 AA; 10018 MW; 194E4AE861F41C74 CRC64;

Query Match
Best Local Similarity 37.5%; Score 3; DB 6; Length 91;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 syxg 8
Db 71 SYXG 74

RESULT 4
O91947 PRELIMINARY; PRT; 131 AA.
ID O91947
AC O91947
DT 04-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE EYA3 HOMOLOGUE (FRAGMENT).

```

```

OS Oryzias latipes (Medaka fish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Belontiiformes; Adrianiichthyidae; Oryziinae; Oryzias.
OX NCBI_TaxID=8090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HNT; TISSUE=WHOLE EMBRYO;
RA Morita Y., Mitani H., Naruse K.;
RL "Oryzias latipes eya3 homologue partial sequence.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB032898; BAA97579.1; -.
FT NON_TER 1
FT NON_TER 131
FT SEQUENCE 131 AA; 14225 MW; A43D41A25564E9C3 CRC64;

Query Match
Best Local Similarity 37.5%; Score 3; DB 13; Length 131;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 syxg 8
Db 67 SYXG 70

RESULT 5
O9GWB2 PRELIMINARY; PRT; 136 AA.
ID O9GWB2
AC O9GWB2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PROBABLE SINGLE STRAND-SPECIFIC NUCLEASE (FRAGMENT).
GN LM12.981.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,
RL Oliver K.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL390114; CAC02622.1; -.
FT NON_TER 1
FT NON_TER 136
FT SEQUENCE 136 AA; 14910 MW; 018F661A2600F9BC CRC64;

Query Match
Best Local Similarity 37.5%; Score 3; DB 5; Length 136;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 syxg 8
Db 52 SYXG 55

RESULT 6
O27953 PRELIMINARY; PRT; 141 AA.
ID O27953
AC O27953;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE B-CASEIN (FRAGMENT).
OS Balaeoptera physalus (Pinback whale) (Common orqual).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
OC Balaeoptera; Balaeoptera.
OX NCBI_TaxID=9770;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RA Gately J., Hayashi C., Cronin M., Arclander P.;
RL Mol. Biol. Evol. 0:0-0(0).
DR EMBL: U53900; AAB08405.1; -.
DR InterPro: IPR001588; Casein.
DR Pfam: PF00363; caseins; 1.
FT NON_TER 1
FT NON_TER 141
SQ SEQUENCE 141 AA; 15822 MW; 7C3EDEE320034513 CRC64;

Query Match 37.5%; Score 3; DB 6; Length 141;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
   ||||
DB 13 SYXG 16

RESULT 7
ID 09N274 PRELIMINARY; PRT: 226 AA.
AC 09N274;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE EXTRACELLULAR GLUTATHIONE PEROXIDASE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Carnivora; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Comhair S.A.A., Thomassen M.J., Erzurum S.C.;
RT "Differential induction of nitric oxide synthase 2 and extracellular
RT glutathione peroxidase in airways of healthy individuals exposed to
RT 100% O2 or cigarette smoke.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF217787; AAF3005.1; -.
DR InterPro: IPR000889; Glut_peroxidase.
DR Pfam: PF00255; GSHPx; 1.
DR PRINTS: PR01011; GLUTPROXDASE.
DR PROSITE: PS00763; GLUTATHIONE_PEROXID_2; 1.
KM Peroxidase.
SQ SEQUENCE 226 AA; 25512 MW; 079B970F7C0651A4 CRC64;

Query Match 37.5%; Score 3; DB 4; Length 226;
Best Local Similarity 100.0%; Pred. No. 2.4e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
   ||||
DB 71 SYXG 74

RESULT 8
ID 09GYV3 PRELIMINARY; PRT: 485 AA.
AC 09GYV3;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE VARIAM SURFACE GLYCOPROTEIN.
OS Trypanosoma brucei.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5691;
RN [1]
RP SEQUENCE FROM N.A.
RA Wickstead B., Ersfeld K., Gull K.;
```

```

RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF294807; AAG03079.1; -.
DR InterPro: IPR001812; Trypan_glycop.
DR Pfam: PF00913; Trypan_glycop; 1.
SQ SEQUENCE 485 AA; 52687 MW; 9D5C330AF1167836 CRC64;

Query Match 37.5%; Score 3; DB 5; Length 485;
Best Local Similarity 100.0%; Pred. No. 4.4e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
   ||||
DB 326 SYXG 329

RESULT 9
ID 09BB52 PRELIMINARY; PRT: 707 AA.
AC 09BB52;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE NADH DEHYDROGENASE F (FRAGMENT).
GN NDHF.
OS Schoenoecephalum cucullatum.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Commelinidae; Inceratae; sedis;
OC Rapateceae; Schoenoecephalum.
OX NCBI_TaxID=142441;
RN [1]
RP SEQUENCE FROM N.A.
RA Givnish T.J., Evans T.M., Zjhra M.L., Patterson T.B., Berry P.E.,
RA Sytma K.J.;
RT "Molecular evolution, adaptive radiation, and geographic
RT diversification in the amphiatlantic family Rapateceae: evidence from
RT ndhf sequences and morphology.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF207634; AAK21842.1; -.
KW Chloroplast.
FT NON_TER 1
FT NON_TER 707
FT NON_TER 707
SQ SEQUENCE 707 AA; 79703 MW; D02C25580E2CB659 CRC64;

Query Match 37.5%; Score 3; DB 8; Length 707;
Best Local Similarity 100.0%; Pred. No. 6e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
   ||||
DB 523 SYXG 526

RESULT 10
ID 095945 PRELIMINARY; PRT: 7 AA.
AC 095945;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE INSIDE INTRON 5 (FRAGMENT).
OS Saccharomyces cerevisiae (Baker's yeast).
OC Mitochondrion.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Bonitz S.G., Coruzzi G., Thalenfeld B., Tzagoloff A., Macino G.;
RA "Assembly of the mitochondrial membrane system. Structure and
```

RT nucleotide sequence of the gene coding for subunit 1 of yeast
RT cytochrome oxidase."
RL J. Biol. Chem. 255:11927-11941(1980).
DR EMBL; V00694; CAA24066.1; -.
KW Mitochondrion.
FT NON_TER
SQ SEQUENCE 7 AA; 859 MW; 75B7232362CDC460 CRC64;

Query Match 25.0%; Score 2; DB 8; Length 7;
Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wk 2
DB 4 wk 5

RESULT 11
ID Q15898 PRELIMINARY; PRT; 8 AA.
AC Q15898;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-AUG-1998 (TREMBLrel. 07, last annotation update)
DE (CLONE XPE6A11B) (FRAGMENT).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Lee C.C., Yazdani A., Wehnert M., Bailey J., Couch L., Xiong M.,
RA Coolbaugh M.I., Chinault C.A., Baldini A., Lindsay E.A., Zhao Z.Y.,
RA Caskey C.T.H.;
RL Hum. Mol. Genet. 0:0-0(0).
DR EMBL; L32078; AAA73888.1; -.
FT NON_TER
SQ SEQUENCE 8 AA; 938 MW; 34AA15B0477B45BB CRC64;

Query Match 25.0%; Score 2; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sy 6
DB 2 sy 3

RESULT 12
ID Q34909 PRELIMINARY; PRT; 8 AA.
AC Q34909;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, last annotation update)
DE CYTOCHROME B (FRAGMENT).
OS Locusta migratoria (Migratory Locust).
OC Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
OC Acridomorpha; Acridoidea; Acrididae; Locusta.
OX NCBI_TaxID=7004;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88223478; PubMed=2836084;
RA WeCracken A., Uhlenbusch I., Gellissen G.;
RT "Structure of the cloned locusta migratoria mitochondrial genome:
RT restriction mapping and sequence of its ND-1 (URF-1) gene.";
RL Curr. Genet. 11:625-630(1987).

DR EMBL; X05286; CAA28905.1; -.
KW Mitochondrion.
FT NON_TER
SQ SEQUENCE 8 AA; 1019 MW; F8E33723304B45B6 CRC64;

Query Match 25.0%; Score 2; DB 8; Length 8;
Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sy 6
DB 2 sy 3

RESULT 13
ID Q9S824 PRELIMINARY; PRT; 8 AA.
AC Q9S824;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE PHOTOSYSTEM I REACTION CENTER SUBUNIT IV, PSI-E.
OS Sphacria oleracea (Spinach).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.
OX NCBI_TaxID=3562;
RN [1]
RP SEQUENCE.
RX MEDLINE=92249324; PubMed=1374333;
RA Lagoute B., Vallon O.;
RT "Purification and membrane topology of PSI-D and PSI-E, two subunits
RT of the photosystem I reaction center.";
RL Eur. J. Biochem. 205:1175-1185(1992).
SQ SEQUENCE 8 AA; 1082 MW; 2145BB1324069044 CRC64;

Query Match 25.0%; Score 2; DB 10; Length 8;
Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sy 6
DB 7 sy 8

RESULT 14
ID Q62721 PRELIMINARY; PRT; 8 AA.
AC Q62721;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, last annotation update)
DE PROHIBITIN (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER;
RX MEDLINE=95331633; PubMed=7607556;
RA Altus M.S., Wood C.W., Stewart D.A., Roskams A.I., Friedman V.,
RA Henderson T., Owens G.A., Danner D.B., Jupe E.R., Dell'Orco R.T.,
RA McClung J.K.;
RT "Regions of evolutionary conservation between the rat and human
RT prohibitin-encoding genes.";
RL Gene 158:291-294(1995).
DR EMBL; U17178; AAA86692.1; -.
FT NON_TER
SQ SEQUENCE 8 AA; 1150 MW; EFD3237B05A41376 CRC64;

Query Match 25.0%; Score 2; DB 11; Length 8;
Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
11
DB 6 wk 7

RESULT 15

O9UQW0
ID O9UQW0 PRELIMINARY; PRT: 9 AA.

AC O9UQW0;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, last annotation update)
DE PROLACTIN PRECURSOR (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84183507; PubMed=6325171;
RA Truong A.T., Duez C., Belayew A., Renard A., Pictet R., Bell G.I.,
Martial J.A.;
RT "Isolation and characterization of the human prolactin gene.";
RL EMBL J. 3:429-437(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93076813; PubMed=1332868;
RA Peers B., Naida A.M., Monget P., Voz M.L., Belayew A., Martial J.A.;
RT "Binding of a 100-kDa ubiquitous factor to the human prolactin
RT promoter is required for its basal and hormone-regulated activity.";
RL Eur. J. Biochem. 210:53-58(1992).
DR EMBL; X00368; CAA25108.1; -.
KW Signal.
FT SIGNAL 1 8 POTENTIAL.
FT NON_TER 9
SQ SEQUENCE 9 AA; 1060 MW; 0A1A677588733054 CRC64;

Query Match 25.0%; Score 2; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
11
DB 8 wk 9

RESULT 16

O9UQW0
ID O9UQW0 PRELIMINARY; PRT: 9 AA.

AC O9UQW0;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, last annotation update)
DE KIT PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94061059; PubMed=7694728;
RA Spitz R.A., Holmes S.A., Berg S.Z., Nordlund J.J., Fukui K.;
RT "A recurrent deletion in the KIT (mast/stem cell growth factor
RT receptor) proto-oncogene is a frequent cause of human piebaldism.";
RL Hum. Mol. Genet. 2:1499-1500(1993).
DR EMBL; S67686; AADI3996.1; -.

FT NON_TER 1 1
SQ SEQUENCE 9 AA; 1182 MW; 0BC504032361B5AB CRC64;

Query Match 25.0%; Score 2; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
11
DB 5 wk 6

RESULT 17

O9HYF9
ID O9HYF9 PRELIMINARY; PRT: 9 AA.

AC O9HYF9;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE CYTOKERATIN 19 (FRAGMENT).
GN K19.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kagaya M., Kaneko S., Ohno H., Inamura K., Kobayashi K.;
RT "Promoter activity and protein binding sites of the regulatory
RT sequences of the human cytokeratin 19 gene.";
RL Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB045973; BAB40770.1; -.
KW Keratin.
FT NON_TER 9
SQ SEQUENCE 9 AA; 1122 MW; 9E9FC41B45AB45A1 CRC64;

Query Match 25.0%; Score 2; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
11
DB 3 sy 4

RESULT 18

O28093
ID O28093 PRELIMINARY; PRT: 9 AA.

AC O28093;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-NOV-1996 (TREMBLrel. 08, last annotation update)
DE CYTOKERATIN IV GENE UPSTREAM REGION (FRAGMENT).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovine; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89231609; PubMed=2469572;
RA Blessing M., Jorcano J.L., Franke W.W.;
RT "Enhancer elements directing cell-type-specific expression of
RT cytokeratin genes and changes of the epithelial cytoskeleton by
RT transfections of hybrid cytokeratin genes.";
RL EMBL J. 8:117-126(1989).
DR EMBL; X14478; CAA32640.1; -.
KW Keratin; Intermediate filament.
FT NON_TER 9
SQ SEQUENCE 9 AA; 1044 MW; 819A22D1B5B32B45 CRC64;

Query Match 25.0%; Score 2; DB 6; Length 9;
Best Local Similarity 100.0%; Pred.No.4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
||
DB 2 SY 3

RESULT 19

09QWTO PRELIMINARY; PRT; 9 AA.
AC 09QWTO; (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE PROTEINASE 3 (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SV129 D3;
RA Lutz P.G., Moog-Lutz C., Houzel-Charavel A., Cayre Y.E.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ007030; CA07429.1; -.
FT NON_TER
SQ SEQUENCE 9 AA; 937 MW; C91E75A77B45B87D CRC64;

Query Match 25.0%; Score 2; DB 11; Length 9;
Best Local Similarity 100.0%; Pred.No.4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
||
DB 4 SY 5

RESULT 20

09IA14 PRELIMINARY; PRT; 9 AA.
AC 09IA14; (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE FIBROBLAST GROWTH FACTOR HOMOLOGOUS FACTOR 2 ISOFORM 1U (FRAGMENT).
FH F-2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20112823; PubMed=10644718;
RA Munoz-Sanjuan I., Smilwood P.M., Nathans J.;
RT "Isoform Diversity among Fibroblast Growth Factor Homologous Factors
is Generated by Alternative Promoter Usage and Differential
Splicing";
RL J. Biol. Chem. 275:2589-2597 (2000).
DR EMBL; AF199609; AAF31396.1; -.
FT NON_TER
SQ SEQUENCE 9 AA; 1068 MW; DF9245B32407272D CRC64;

Query Match 25.0%; Score 2; DB 13; Length 9;
Best Local Similarity 100.0%; Pred.No.4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy, 6

DB 7 SY 8
||

RESULT 21

09XMB4 PRELIMINARY; PRT; 10 AA.
AC 09XMB4; (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DE 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT 3 (FRAGMENT).
GN NAD3.
OS Aegilops tauschii (Patropyrum tauschii).
OC Microchondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Triticeae; Aegilops.
OX NCBI_TaxID=37682;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K029;
RA Tsukamoto N., Asakura N., Takumi S., Mori N., Nakamura C.;
RT "The presence of paternal sub-genomic mitochondrial DNA copies in the
nucleus-cytoplasm hybrids of tetraploid wheat with the cytoplasm of
Aegilops squarrosa";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF142479; AAD37355.1; -.
FT NON_TER
SQ SEQUENCE 10 AA; 1233 MW; 5F9A1B5BDD8403 CRC64;

Query Match 25.0%; Score 2; DB 8; Length 10;
Best Local Similarity 100.0%; Pred.No.4.1e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 WK 2
||
DB 2 WK 3

RESULT 22

09UC46 PRELIMINARY; PRT; 11 AA.
AC 09UC46; (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE NEUTROPHIL INHIBITOR PEPTIDE, NIP=POLYMORPHONUCLEAR NEUTROPHIL
DE INHIBITOR PEPTIDE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=96326114; PubMed=8703476;
RA Cooper J.A., Jr., Culbreth R.R.;
RT "Characterization of a neutrophil inhibitor peptide harvested from
human bronchial lavage: homology to influenza A nucleoprotein";
RL Am. J. Respir. Cell Mol. Biol. 15:207-215 (1996).
SQ SEQUENCE 11 AA; 1262 MW; 951A1C3279C9DB45 CRC64;

Query Match 25.0%; Score 2; DB 4; Length 11;
Best Local Similarity 100.0%; Pred.No.4.5e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
||
DB 4 SY 5


```
RESULT 23
ID 077876 PRELIMINARY: PRT: 11 AA.
AC 077876:
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DE MHC CLASS II B LOCUS 1 (FRAGMENT).
OS Oreochromis niloticus (Nile tilapia) (Tilapia nilotica).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Labroidel;
OC Cichlidae; Oreochromis.
NCBI_TaxID=8128;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98315113; PubMed=9649539;
RA Malaga-Trillo E., McAndrew B., Vincek V., Zaleska-Rutczynska Z.,
RA Sueltmann H., Figueroa F., Klein J.;
RT "Linkage relationships and haplotype polymorphism among cichlid Mhc
RT class II B loci."
RL Genetics 149:1527-1547(1998).
DR EMBL: AF049985; AAC41324.1; -.
KW MHC.
FT NON_TER 1 1
FT NON_TER 11 11
SQ SEQUENCE 11 AA: 1178 MW: 9AC131FAB2D2DBA5 CRC64:

Query Match 25.0%; Score 2; DB 7; Length 11;
Best Local Similarity 100.0%; Pred. No. 4.5e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
DB 4 sy 5

RESULT 24
ID 077883 PRELIMINARY: PRT: 11 AA.
AC 077883:
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DE MHC CLASS II B LOCUS 4 (FRAGMENT).
OS Oreochromis niloticus (Nile tilapia) (Tilapia nilotica).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Labroidel;
OC Cichlidae; Oreochromis.
NCBI_TaxID=8128;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98315113; PubMed=9649539;
RA Malaga-Trillo E., McAndrew B., Vincek V., Zaleska-Rutczynska Z.,
RA Sueltmann H., Figueroa F., Klein J.;
RT "Linkage relationships and haplotype polymorphism among cichlid Mhc
RT class II B loci."
RL Genetics 149:1527-1547(1998).
DR EMBL: AF049992; AAC41331.1; -.
KW MHC.
FT NON_TER 1 1
FT NON_TER 11 11
SQ SEQUENCE 11 AA: 1288 MW: 87F42A0FB2D5ABA5 CRC64:
```

```
Query Match 25.0%; Score 2; DB 7; Length 11;
Best Local Similarity 100.0%; Pred. No. 4.5e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
```

```
DB 4 sy 5

RESULT 25
ID 077886 PRELIMINARY: PRT: 11 AA.
AC 077886:
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DE MHC CLASS II B LOCUS 4 (FRAGMENT).
OS Oreochromis niloticus (Nile tilapia) (Tilapia nilotica).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Labroidel;
OC Cichlidae; Oreochromis.
NCBI_TaxID=8128;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98315113; PubMed=9649539;
RA Malaga-Trillo E., McAndrew B., Vincek V., Zaleska-Rutczynska Z.,
RA Sueltmann H., Figueroa F., Klein J.;
RT "Linkage relationships and haplotype polymorphism among cichlid Mhc
RT class II B loci."
RL Genetics 149:1527-1547(1998).
DR EMBL: AF049995; AAC41334.1; -.
KW MHC.
FT NON_TER 1 1
FT NON_TER 11 11
SQ SEQUENCE 11 AA: 1343 MW: 87F42D9F52D41B45 CRC64:
```

```
Query Match 25.0%; Score 2; DB 7; Length 11;
Best Local Similarity 100.0%; Pred. No. 4.5e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
DB 4 sy 5
```

Search completed: January 14, 2002, 08:07:30
Job time: 760 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:44:25 ; Search time 81.39 Seconds
(without alignments)
7.281 Million cell updates/sec

Title: 09-185908-1A
Perfect score: 8
Sequence: 1 wksxyxg 8

Scoring table: OLIGO
Gapop 60.0 , Gapect 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

Database :

- 1: /SID52/gcgdata/geneseq/geneseqp/AA1980.DAT.*
- 2: /SID52/gcgdata/geneseq/geneseqp/AA1981.DAT.*
- 3: /SID52/gcgdata/geneseq/geneseqp/AA1982.DAT.*
- 4: /SID52/gcgdata/geneseq/geneseqp/AA1983.DAT.*
- 5: /SID52/gcgdata/geneseq/geneseqp/AA1984.DAT.*
- 6: /SID52/gcgdata/geneseq/geneseqp/AA1985.DAT.*
- 7: /SID52/gcgdata/geneseq/geneseqp/AA1986.DAT.*
- 8: /SID52/gcgdata/geneseq/geneseqp/AA1987.DAT.*
- 9: /SID52/gcgdata/geneseq/geneseqp/AA1988.DAT.*
- 10: /SID52/gcgdata/geneseq/geneseqp/AA1989.DAT.*
- 11: /SID52/gcgdata/geneseq/geneseqp/AA1990.DAT.*
- 12: /SID52/gcgdata/geneseq/geneseqp/AA1991.DAT.*
- 13: /SID52/gcgdata/geneseq/geneseqp/AA1992.DAT.*
- 14: /SID52/gcgdata/geneseq/geneseqp/AA1993.DAT.*
- 15: /SID52/gcgdata/geneseq/geneseqp/AA1994.DAT.*
- 16: /SID52/gcgdata/geneseq/geneseqp/AA1995.DAT.*
- 17: /SID52/gcgdata/geneseq/geneseqp/AA1996.DAT.*
- 18: /SID52/gcgdata/geneseq/geneseqp/AA1997.DAT.*
- 19: /SID52/gcgdata/geneseq/geneseqp/AA1998.DAT.*
- 20: /SID52/gcgdata/geneseq/geneseqp/AA1999.DAT.*
- 21: /SID52/gcgdata/geneseq/geneseqp/AA2000.DAT.*
- 22: /SID52/gcgdata/geneseq/geneseqp/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37.5	24	20	AAW74457	Fibroblast growth
2	37.5	24	21	AAV90893	Peptide #2273 enco
3	37.5	34	22	AAV15839	Peptide #2387 enco
4	37.5	34	22	AAV28350	Peptide #2259 enco
5	37.5	34	22	AAV03577	Peptide #2259 enco
6	37.5	53	11	AAV08007	Modified murine ep
7	37.5	59	21	AAV00560	Human secreted pro
8	37.5	59	21	AAV65150	Human 5' EST relat
9	37.5	72	22	AAV77776	Human colon cancer
10	37.5	90	21	AAV28199	Arabidopsis thalia
11	37.5	96	21	AAV28198	Arabidopsis thalia

12	3	37.5	104	14	AAV38596	Human lambda light
13	3	37.5	104	19	AAV58496	Human lambda light
14	3	37.5	105	21	AAV02234	Human secreted pro
15	3	37.5	129	22	AAV57533	Human colon cancer
16	3	37.5	164	21	AAV28197	Arabidopsis thalia
17	3	37.5	226	16	AAV72680	Human plasma gluta
18	3	37.5	226	21	AAV22828	Human plasma gluta
19	3	37.5	230	18	AAV08430	Rapamycin-dependen
20	3	37.5	329	21	AAV08728	Amino acid sequenc
21	3	37.5	533	15	AAV4693	Diphtheria toxin (
22	3	37.5	684	17	AAV17581	Thermoaerobacter
23	3	37.5	684	17	AAV17585	Thermoaerobacter
24	3	37.5	684	17	AAV17589	Thermoaerobacter
25	3	37.5	685	17	AAV17582	Thermoaerobacter
26	3	37.5	685	17	AAV17586	Thermoaerobacter
27	3	37.5	685	17	AAV17590	Thermoaerobacter
28	3	37.5	685	17	AAV17590	ACE inhibitor SP3.
29	2	25.0	3	15	AAV42556	Sequence of crypt
30	2	25.0	4	2	AAV10154	Generic enkephalin
31	2	25.0	4	2	AAV10370	Enkephalin-like an
32	2	25.0	4	2	AAV10372	Enkephalin-like an
33	2	25.0	4	2	AAV10373	Enkephalin-like an
34	2	25.0	4	2	AAV10375	Enkephalin-like an
35	2	25.0	4	2	AAV10386	Generic analgesic
36	2	25.0	4	2	AAV10599	N-adamantane tetra
37	2	25.0	4	2	AAV10433	Analgesic tetrapep
38	2	25.0	4	2	AAV10401	Analgesic tetrapep
39	2	25.0	4	2	AAV10620	Analgesic tetrapep
40	2	25.0	4	2	AAV10625	Analgesic and neur
41	2	25.0	4	3	AAV20210	Analgesic and neur
42	2	25.0	4	3	AAV20210	Sequence of enkeph
43	2	25.0	4	5	AAV40339	Sequence of Cyclo(
44	2	25.0	4	6	AAV50545	Sequence of Cyclo(
45	2	25.0	4	7	AAV51359	Organ specific neo
46	2	25.0	4	9	AAV81591	Cyclic enkephalin
47	2	25.0	4	11	AAV04767	Enkephalin analogu
48	2	25.0	4	11	AAV07020	Enkephalin analogu
49	2	25.0	4	12	AAV12507	Enkephalin analogu
50	2	25.0	4	12	AAV12508	5 amino acid resid

ALIGNMENTS

RESULT 1	AAV74457	standard; peptide; 24 AA.
ID	AAV74457	
XX	AAV74457	
DE	12-MAY-1999 (first entry)	
XX	Fibroblast growth factor 3 receptor mutational cassette.	
XX	Agonist identification; orphan receptor; constitutively active OR;	
KW	Graves' disease; thyroid adenoma; hypertension; cardiomyopathy;	
KW	Schizophrenia; Kaposi's sarcoma; fibroblast growth factor receptor;	
KW	adenylate cyclase constitutive activator; thyrotropin receptor;	
KW	thyrotropin stimulating hormone; beta-adrenergic receptor.	
XX		
OS	Homo sapiens.	
XX		
FT	Key	Location/Qualifiers
FT	Misc-difference 9	/label= Glu, Asp, Gln, His, Lys
XX		
XX	W09846995-A1.	
XX		
PD	22-OCT-1998.	
XX		
PF	14-APR-1998; 98MO-US07496.	
XX		
PR	14-APR-1997; 97US-0839449.	

XX (BEHA/) BEHAN D P.
PA (CHAL/) CHALMERS D T.
XX Behan DP, Chalmers DT;
PI WPI: 1999-105468/09.
XX
XX Identifying agonists of orphan receptors from their effect on the
PT constitutively active receptor - particularly therapeutically active
PT inverse agonists at G protein coupled receptors, without requiring
PT knowledge of endogenous ligand or receptor function
XX
PS Disclosure: Page 47; 114pp; English.
XX
CC This sequence is a fragment of a fibroblast growth factor 3 receptor
CC mutational cassette. The invention relates to a method for the
CC identification of candidate compounds as agonists, including inverse or
CC partial, of an orphan receptor (OR), which comprises: (1) applying test
CC compound to constitutively active OR; and (ii) measuring its effect on
CC OR. The method is particularly used to identify inverse agonists of
CC G protein-coupled OR, i.e. potential therapeutic agents for treating
CC conditions in which constitutively active OR are implicated (e.g. Graves'
CC disease, thyroid adenoma, hypertension, cardiomyopathy, schizophrenia,
CC major depression, Kaposi's sarcoma and many others tabulated). It is
CC based on identification of agents that reduce receptor activation, rather
CC than compounds that antagonise the normal ligand. Once identified,
CC (inverse) agonists can be used to study OR function. The method does
CC not require knowledge of the endogenous receptor ligand or receptor
CC function, and identifies directly compounds that inhibit the activated
CC receptor, i.e. able to block both ligand-dependent and -independent
CC activation, rather than only the ligand-dependent process, as is the
CC case with compounds identified by ligand-dependent assays. It should
CC accelerate drug discovery at a wide range of OR and since activated
CC receptors have a greater response to the agents, potential drugs are more
CC likely to be detected.
CC
XX Sequence 24 AA:
SQ

Query Match 37.5%; Score 3; DB 20; Length 24;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
|||
Db 7 syxg 10

RESULT 2
ID AAY90893 standard; peptide; 24 AA.
AC AAY90893;
XX
XX 30-AUG-2000 (first entry)
DT
XX
DE Fibroblast growth factor 3 receptor mutational cassette #1.
XX
KW Identification: modulator; cell surface membrane receptor; treatment;
KW orphan receptor; antithyroid; antidiabetic; neuroleptic; antidepressant;
KW cytosolic; G protein-coupled receptor agonist.
XX
XX Synthetic.
OS
XX
FH Key Location/Qualifiers
FT Misc-difference 9 /Label= Glu, Asp, Gln, His, Lys
XX
XX WO200021987-A2.
XX
XX 20-APR-2000.
XX

PF 12-OCT-1999; 99WO-US23935.
XX
XX 13-OCT-1998; 98US-0170496.
XX
PA (AREN-) ARENA PHARM INC.
XX
PI Behan DP, Chalmers DT;
XX
XX WPI: 2000-317935/27.
DR
XX
XX Identifying compounds with inverse agonist activity to orphan receptors
PT useful for treating e.g. Graves' disease, and schizophrenia, involves
PT contacting candidate compounds with constitutively activated receptors
PT
XX
PS Disclosure: Page 47; 110pp; English.
XX
XX The present invention describes a method for directly identifying a
CC compound having inverse agonist activity, partial agonist activity or
CC agonist activity to a constitutively active orphan receptor (ORR). The
CC method comprises determining the efficacy of the compound by contacting
CC it with the ORR. A compound identified by the above method having
CC inverse agonist activity to ORR is useful for the treatment of diseases
CC characterised by constitutive activation of the receptor e.g. Graves'
CC disease, male precocious puberty, Jansen's disease, retinitis pigmentosa,
CC hypoparathyroidism, neuropsychiatric diseases, schizophrenia, major
CC depression, and cancerous growth in Kaposi's sarcoma. The method can
CC identify (i) directly without prior knowledge or use of receptor ligands
CC and is useful for accelerating drug discovery at a broad range of ORR.
CC The present sequence represents a fibroblast growth factor 3 receptor
CC mutational cassette, which is used in the exemplification of
CC the present invention.
XX
XX Sequence 24 AA:
SQ

Query Match 37.5%; Score 3; DB 21; Length 24;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
|||
Db 7 syxg 10

RESULT 3
ID AAM15839 standard; Protein; 34 AA.
AC AAM15839;
XX
XX 12-OCT-2001 (first entry)
DT
XX
DE Peptide #2273 encoded by probe for measuring cervical gene expression.
XX
XX Probe: human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer.
XX
XX Homo sapiens.
OS
XX
XX WO200157278-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00670.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
PR

XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI: 2001-488901/53.
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
XX Claim 27: SEQ ID No 20665; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENPs: see AA110068-AA128459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human Hela cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 34 AA:
SO
Query Match 37.5%; Score 3; DB 22; Length 34;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 syxg 8
|||||
Db 18 syxg 21
RESULT 4
AAAM28350
ID AAAM28350 standard; Protein: 34 AA.
XX
XX AAAM28350;
XX
DT 18-OCT-2001 (first entry)
XX
XX Peptide #2387 encoded by probe for measuring placental gene expression.
XX
XX Probe: microarray: human; placenta; antenatal diagnosis;
XX genetic disorder.
XX
XX Homo sapiens.
XX
XX MO200157272-A2.
XX
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI: 2001-48897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -

XX
XX Claim 27: SEQ ID No 28619; 654pp; English.
XX
XX
CC The present invention relates to single exon nucleic acid probes (SENPs:
CC see AA131315-AA157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
XX
XX
XX Sequence 34 AA:
SQ
Query Match 37.5%; Score 3; DB 22; Length 34;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 syxg 8
|||||
Db 18 syxg 21
RESULT 5
AAAM03577
ID AAAM03577 standard; Protein: 34 AA.
XX
XX AAAM03577;
XX
DT 09-OCT-2001 (first entry)
XX
XX
XX Peptide #2259 encoded by probe for measuring breast gene expression.
XX
XX
XX Probe: human; breast disease; breast cancer; development disorder;
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX Homo sapiens.
XX
XX MO200157270-A2.
XX
XX
XX 09-AUG-2001.
XX
XX 29-JAN-2001; 2001WO-US00661.
XX
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI: 2001-476286/51.
XX
XX
XX Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
XX
XX
XX Claim 27: SEQ ID No 12317; 322pp; English.
XX
XX
CC The present invention relates to novel single exon nucleic acid probes
CC (see AA100010-AA110067). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for measuring human gene expression in
CC a human breast sample, where the probe hybridises at high stringency to a
CC nucleic acid expressed in the human breast. The probes are useful for
CC predicting, diagnosing, grading, staging, monitoring and prognosing
CC diseases of the human breast, particularly those diseases with polygenic
CC aetiology. The diseases include: breast cancer; disorders of development,
CC inflammatory diseases of the breast; fibrocystic changes; proliferative
CC breast disease and non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 34 AA:

Query Match 37.5%; Score 3; DB 22; Length 34;

Best Local Similarity 100.0%; Pred. No. 4.4e+02; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
 ||||
 Db 18 syxg 21

RESULT 6

AA08007
 ID AAR08007 standard; protein; 53 AA.

XX AAR08007;

XX 25-FEB-1991 (first entry)

XX Modified murine epidermal growth factor.

KW Modified murine epidermal growth factor; stability; storage;

KW epithelial wounds; gastric acid secretion.

XX Key Location/Qualifiers

FT Misc-difference 11..11

FT /label= E, N, Q, A, K

XX W09013570-A.

XX 15-NOV-1990.

XX 09-MAY-1990; 90WO-US02600.

XX 12-MAY-1989; 89US-0351773.

XX (CHIR-) CHIRON CORP.

XX Nascimento CG, Medina-Selby A;

XX WPI: 1990-361427/48.

XX Human epidermal growth factor - is substituted at position 11 for
 PT greater stability and improved storage life.

XX Claim 9; Page 25; 32pp; English.

XX The human EGF is used to treat oversecretion of gastric acid or an
 CC epithelial wound. EGF is modified to increase its chemical
 CC stability. Its storage life is improved without diminishing its
 CC biological activity. The proteins may be prepared by traditional
 CC chemical or recombinant means.
 CC See also AAR08004.

XX Sequence 53 AA:

Query Match 37.5%; Score 3; DB 11; Length 53;

Best Local Similarity 100.0%; Pred. No. 5.8e+02; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
 ||||
 Db 9 syxg 12

RESULT 7

AA00560
 ID AAC00560 standard; protein; 59 AA.

XX AAC00560;

XX 06-OCT-2000 (first entry)

XX Human secreted protein, SEQ ID NO: 4641.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI: 2000-500381/45.

XX N-PSDB; AAC00566.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 13; SEQ ID 4641; 71pp + CD-ROM; English.

XX The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30
 CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (UTR) of the mRNA because they are often obtained
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 CC those cases where longer cDNA sequences have been obtained, the full 5'
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 CC ends and can therefore be used to obtain full length cDNAs and genomic
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.

XX Sequence 59 AA:

Query Match 37.5%; Score 3; DB 21; Length 59;

Best Local Similarity 100.0%; Pred. No. 6.2e+02; Mismatches 0; Indels 0; Gaps 0;

OY 2 kxxy 6
 ||||
 Db 48 kxxy 52

RESULT 8

AA065150
 ID AAY65150 standard; protein; 59 AA.

XX AAY65150;

XX 01-FEB-2000 (first entry)

XX Human 5' EST related polypeptide SEQ ID NO:1311.

XX Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
 KW gene therapy; chromosome mapping; upstream regulatory sequence;
 KW forensic; location; development; protein synthesis; stability;
 KW regulation; identification.

XX	Homo sapiens.
OS	
XX	
PN	M09953051-A2.
PD	21-OCT-1999.
XX	
PF	09-APR-1999; 99WO-1B00712.
XX	
PR	09-APR-1998; 98US-0057719.
PR	28-APR-1998; 98US-0069047.
XX	
PA	(GEST) GENSET.
XX	
PI	Dumas MLine Edwards J, Duclert A, Giordano J;
XX	
DR	WPI: 2000-038446/03.
DR	N-PSDB; AA242764.
XX	
PT	Novel secreted protein 5' expressed sequence tag sequences used in
PT	diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX	
PS	Claim 3; Page 748; 837pp; English.
XX	
CC	AA242265 to AA243075 represent novel 5' expressed sequence tag (EST)
CC	sequences, corresponding to human secreted proteins. AAY64651 to
CC	AA243052. The 5' ESTs can be used for producing secreted human gene
CC	products. They can be used to identify and isolate 5' untranslated
CC	regions (UTRs) and upstream regulatory regions which control the
CC	location, development stage, rate, and quantity of protein synthesis, as
CC	well as stability of mRNA. The ESTs are also useful as probes for
CC	chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC	also be used in forensic procedures to identify individuals, or in
CC	diagnostic procedures to identify individuals having genetic diseases
CC	resulting from abnormal gene expression. The products may also be used in
CC	gene therapy protocols. The nucleic acids encoding signal peptides can be
CC	used for directing extracellular secretion of a polypeptide or the
CC	insertion of a polypeptide into a membrane, or importing a polypeptide
CC	into a cell. The proteins encoded by the EST sequences may be useful in
CC	treating a variety of human conditions. Secreted proteins have
CC	therapeutic value, and the identification of new secreted proteins is
CC	valuable. AA242249 to AA242264 and AAY64644 to AAY64650 represent
CC	sequences used in the exemplification of the present invention.
XX	
SQ	Sequence 59 AA:
XX	
Query Match	37.5%; Score 3; DB 21; Length 59;
Best Local Similarity	100.0%; Pred. No. 6.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 2 kxxxy 6	
IIIII	
Db 48 kxxy 52	
XX	
RESULT 9	
AAG77776	
ID AAG77776 standard; Protein: 72 AA.	
XX	
AAG77776;	
XX	
DT 03-SEP-2001 (first entry)	
XX	
DE Human colon cancer antigen protein SEQ ID NO:8542.	
XX	
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;	
KW colorectal carcinoma.	
XX	
OS Homo sapiens.	
XX	
NN M0200123920-A2.	

[illegible]

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.

PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158332.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 37.5%; Score 3; DB 21; Length 90;
Best Local Similarity 100.0%; Pred. No. 8.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
||||
DB 38 syxg 41

RESULT 11
AAC28198 AAC28198 strand; Protein; 96 AA.

XX AAC28198;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 33327.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 23-MAR-1999; 99US-0123548.

PR 25-MAR-1999; 99US-0125788.

PR 29-MAR-1999; 99US-0126264.

PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135333.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.

PR 13-DEC-1991: 91US-0808464.
XX (XOMA) XOMA CORP.
XX
PI Flashwld DM, Kohn FR, Little RG, Studnicka GM;
XX WPI: 1993-213827/26.
DR
XX Antibodies prepn. used for treatment of auto-immune diseases - by
PT replacement of critical residues to reduce immunogenicity but
PT retain binding affinity, etc.
XX
PS Claim 2: Page 96; 160pp; English.
XX
XX The consensus amino acid sequences for the subgroups of light
CC chains (hK1 - AAR38590, hK3 - NGK, hK2 - GST, hL1 - AAR38591, hL2 -
CC AAR38592, hL3 - AAR38593, hL6 - AAR38594, hK4 - AAR38595, hL4 -
CC AAR38596, and hL5 - AAR38597) and heavy chains (hH3 - AAR38598, hH1 -
CC AAR38599, and hH2 - AAR38600) of human variable domains may be used to
CC prepare, for example, a modified mouse antibody variable domain that
CC retains the affinity of the natural domain for antigen while exhibiting
CC reduced immunogenicity in humans.
CC Unlike other methods of humanisation, which advocate the
CC replacement of entire antibody framework regions with those of human
CC antibodies, this method involves only the introduction of human
CC residues into those positions not critical for antigen binding.
CC This ensures that the binding properties of the modified antibody
CC are not diminished.
XX
XX Sequence 104 AA:
SO
Query Match 37.5%; Score 3; DB 14; Length 104;
Best Local Similarity 100.0%; Pred. No. 8.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 syxg 8
||||
Db 85 syxg 88
RESULT 13
AAW58496
ID AAW58496 standard; protein; 104 AA.
XX
XX AAW58496;
AC
XX
DT 18-AUG-1998 (first entry)
XX
XX Human lambda light chain subgroup 2 consensus sequence hL2.
DE
XX
XX Humanised; human; mouse; CD5; anti-CD5 antibody; immunoglobulin;
KW depletion; cytotoxic; immunconjugate; fusion protein; psoriasis;
KW autoimmune disease; rheumatoid arthritis; type I diabetes.
XX
XX Synthetic.
OS
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH Misc-difference 1 /note= "not specified"
FT
FT Misc-difference 32 /note= "not specified"
FT
FT Misc-difference 87 /note= "not specified"
FT
FT Misc-difference 89 /note= "not specified"
FT
FT Misc-difference 90 /note= "not specified"
FT
FT Misc-difference 91 /note= "not specified"
FT
FT Misc-difference 92 /note= "not specified"
FT
FT Misc-difference 92 /note= "not specified"

XX
PN US5770196-A.
XX
XX 23-JUN-1998.
PD
XX
XX 07-JUN-1995; 95US-0472788.
PF
XX
XX 23-JUN-1993; 93US-0082842.
PR 13-DEC-1991; 91US-0808464.
PR 14-DEC-1992; 92WO-US10906.
PR 07-JUN-1995; 95US-0472788.
XX
XX (XOMA) XOMA CORP.
XX
XX Studnicka GM;
PI
XX WPI: 1998-376744/32.
DR
XX
XX Depletion of CD5-positive cells in vivo - using anti-CD5 antibodies
PT with humanised variable regions
PT
XX
XX Disclosure; Column 55-56; 77pp; English.
PS
XX
XX A method has been developed of depleting CD5+ cells in an animal. The
CC method comprises administering a cytotoxic protein containing a modified
CC immunoglobulin (Ig) variable domain, where the protein is an anti-CD5 Ig
CC molecule or an immunconjugate or fusion protein containing an anti-CD5
CC Ig molecule, and where the modified Ig variable domain comprises at
CC least one of (a) a modified light chain variable region (see AAW58478 or
CC AAW58480), and (b) a modified heavy chain variable region (see AAW58478 or
CC or AAW58481), where AAW58478 and AAW58479 are humanised forms of the H65
CC light and heavy chain variable domains with low risk amino acid
CC substitutions (i.e. low risk of reducing antigen-binding specificity.)
CC and AAW58480 and AAW58481 are humanised forms of the H65 light and heavy
CC chain variable domains with moderate risk amino acid substitutions and
CC are present in humanised H65 antibody h63 (ATCC HB 11206). The method is
CC useful for treating autoimmune diseases, especially systemic lupus
CC erythematosus, rheumatoid arthritis, psoriasis or type I diabetes. The
CC present sequence represents a consensus amino acid sequence for light
CC chain subgroups of human antibody variable domains, from the present
CC invention.
XX
XX Sequence 104 AA:
SQ
Query Match 37.5%; Score 3; DB 19; Length 104;
Best Local Similarity 100.0%; Pred. No. 8.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 syxg 8
||||
Db 85 syxg 88
RESULT 14
AAG02234
ID AAG02234 standard; Protein; 105 AA.
XX
XX AAG02234;
AC
XX
XX 06-OCT-2000 (first entry)
DT
XX
XX Human secreted protein, seq ID NO: 6315.
DE
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.
XX
XX Homo sapiens.
OS
XX
XX EP1033401-A2.
PN
XX
XX 06-SEP-2000.
PD
XX

PF 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GEST) GENSET.
XX
PI Dunas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI: 2000-500381/45.
XX
XX N-PSDB; AAC02240.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 13; SEQ ID 6315; 71pp + CD-ROM; English.
XX
XX The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.
XX
XX Sequence 105 AA:

Query Match 37.5%; Score 3; DB 21; Length 105;
Best Local Similarity 100.0%; Pred. No. 8.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
|||
Db 33 syxg 36

RESULT 15
AAG75733
ID AAG75733 standard; Protein: 129 AA.
XX
XX AAG75733;
XX
XX 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:6497.
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX
XX Homo sapiens.
OS
XX W0200122920-A2.
PN
XX
PD 05-APR-2001.
XX
XX 28-SEP-2000; 2000MO-US26524.
PF
XX 29-SEP-1999; 99US-0157137.
XX
XX 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI
XX WPI: 2001-235357/24.
XX
XX N-PSDB; AAH35138.
DR

XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
PT
XX
XX Claim 11; Page 7974-7975; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patient's own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
XX Sequence 129 AA:

Query Match 37.5%; Score 3; DB 22; Length 129;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
|||
Db 117 syxg 120

RESULT 16
AAG28197
ID AAG28197 standard; Protein: 164 AA.
XX
XX AAG28197;
XX
XX 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 33326.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
OS
XX EP1033405-A2.
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825.
XX
XX 09-MAR-1999; 99US-0123180.
XX
XX 09-MAR-1999; 99US-0123548.
XX
XX 23-MAR-1999; 99US-0125788.
XX
XX 25-MAR-1999; 99US-0126284.
XX
XX 29-MAR-1999; 99US-0126785.
XX
XX 01-APR-1999; 99US-0127462.
XX
XX 06-APR-1999; 99US-0128234.
XX
XX 08-APR-1999; 99US-0128714.
XX
XX 16-APR-1999; 99US-0129845.
XX
XX 19-APR-1999; 99US-0130077.
XX
XX 21-APR-1999; 99US-0130449.
XX
XX 23-APR-1999; 99US-0130510.
PR

PR 23-APR-1999;	99US-0130891.	PR 20-JUL-1999;	99US-0144884.
PR 28-APR-1999;	99US-0131449.	PR 21-JUL-1999;	99US-0144834.
PR 30-APR-1999;	99US-0132048.	PR 21-JUL-1999;	99US-0145086.
PR 04-MAY-1999;	99US-0132407.	PR 21-JUL-1999;	99US-0145088.
PR 05-MAY-1999;	99US-0132484.	PR 22-JUL-1999;	99US-0145085.
PR 06-MAY-1999;	99US-0132485.	PR 22-JUL-1999;	99US-0145087.
PR 06-MAY-1999;	99US-0132486.	PR 22-JUL-1999;	99US-0145089.
PR 07-MAY-1999;	99US-0132487.	PR 22-JUL-1999;	99US-0145192.
PR 11-MAY-1999;	99US-0132863.	PR 23-JUL-1999;	99US-0145145.
PR 14-MAY-1999;	99US-0134256.	PR 23-JUL-1999;	99US-0145218.
PR 14-MAY-1999;	99US-0134219.	PR 23-JUL-1999;	99US-0145224.
PR 14-MAY-1999;	99US-0134221.	PR 26-JUL-1999;	99US-0145276.
PR 14-MAY-1999;	99US-0134370.	PR 27-JUL-1999;	99US-0145913.
PR 18-MAY-1999;	99US-0134370.	PR 27-JUL-1999;	99US-0145918.
PR 18-MAY-1999;	99US-0134768.	PR 27-JUL-1999;	99US-0145919.
PR 19-MAY-1999;	99US-0134941.	PR 28-JUL-1999;	99US-0145951.
PR 20-MAY-1999;	99US-0135124.	PR 02-AUG-1999;	99US-0146386.
PR 21-MAY-1999;	99US-0135353.	PR 02-AUG-1999;	99US-0146388.
PR 24-MAY-1999;	99US-0135629.	PR 02-AUG-1999;	99US-0146389.
PR 25-MAY-1999;	99US-0136021.	PR 03-AUG-1999;	99US-0147038.
PR 27-MAY-1999;	99US-0136392.	PR 04-AUG-1999;	99US-0147204.
PR 28-MAY-1999;	99US-0136782.	PR 04-AUG-1999;	99US-0147302.
PR 01-JUN-1999;	99US-0137222.	PR 05-AUG-1999;	99US-0147302.
PR 03-JUN-1999;	99US-0137528.	PR 05-AUG-1999;	99US-0147260.
PR 04-JUN-1999;	99US-0137502.	PR 06-AUG-1999;	99US-0147303.
PR 07-JUN-1999;	99US-0137724.	PR 06-AUG-1999;	99US-0147416.
PR 08-JUN-1999;	99US-0138094.	PR 09-AUG-1999;	99US-0147493.
PR 10-JUN-1999;	99US-0138540.	PR 09-AUG-1999;	99US-0147935.
PR 10-JUN-1999;	99US-0138847.	PR 10-AUG-1999;	99US-0148171.
PR 14-JUN-1999;	99US-0139119.	PR 11-AUG-1999;	99US-0148319.
PR 16-JUN-1999;	99US-0139452.	PR 12-AUG-1999;	99US-0148341.
PR 16-JUN-1999;	99US-0139453.	PR 13-AUG-1999;	99US-0148365.
PR 17-JUN-1999;	99US-0139492.	PR 13-AUG-1999;	99US-0148684.
PR 18-JUN-1999;	99US-0139454.	PR 16-AUG-1999;	99US-0149368.
PR 18-JUN-1999;	99US-0139455.	PR 17-AUG-1999;	99US-0149175.
PR 18-JUN-1999;	99US-0139456.	PR 18-AUG-1999;	99US-0149426.
PR 18-JUN-1999;	99US-0139457.	PR 20-AUG-1999;	99US-0149722.
PR 18-JUN-1999;	99US-0139458.	PR 20-AUG-1999;	99US-0149723.
PR 18-JUN-1999;	99US-0139459.	PR 20-AUG-1999;	99US-0149929.
PR 18-JUN-1999;	99US-0139460.	PR 23-AUG-1999;	99US-0149902.
PR 18-JUN-1999;	99US-0139461.	PR 23-AUG-1999;	99US-0149930.
PR 18-JUN-1999;	99US-0139462.	PR 25-AUG-1999;	99US-0150566.
PR 18-JUN-1999;	99US-0139463.	PR 26-AUG-1999;	99US-0150884.
PR 18-JUN-1999;	99US-0139750.	PR 27-AUG-1999;	99US-0151065.
PR 18-JUN-1999;	99US-0138763.	PR 27-AUG-1999;	99US-0151066.
PR 21-JUN-1999;	99US-0139817.	PR 27-AUG-1999;	99US-0151080.
PR 22-JUN-1999;	99US-0139899.	PR 30-AUG-1999;	99US-0151303.
PR 23-JUN-1999;	99US-0140353.	PR 31-AUG-1999;	99US-0151438.
PR 23-JUN-1999;	99US-0140354.	PR 01-SEP-1999;	99US-0151930.
PR 24-JUN-1999;	99US-0140695.	PR 07-SEP-1999;	99US-0152363.
PR 28-JUN-1999;	99US-0140823.	PR 10-SEP-1999;	99US-0153070.
PR 29-JUN-1999;	99US-0140991.	PR 13-SEP-1999;	99US-0153758.
PR 30-JUN-1999;	99US-0141287.	PR 15-SEP-1999;	99US-0154018.
PR 01-JUL-1999;	99US-0141842.	PR 16-SEP-1999;	99US-0154039.
PR 01-JUL-1999;	99US-0142154.	PR 20-SEP-1999;	99US-0154779.
PR 02-JUL-1999;	99US-0142055.	PR 22-SEP-1999;	99US-0155139.
PR 06-JUL-1999;	99US-0142390.	PR 23-SEP-1999;	99US-0155486.
PR 08-JUL-1999;	99US-0142803.	PR 24-SEP-1999;	99US-0155659.
PR 09-JUL-1999;	99US-0142920.	PR 28-SEP-1999;	99US-0156458.
PR 12-JUL-1999;	99US-0142977.	PR 29-SEP-1999;	99US-0156596.
PR 13-JUL-1999;	99US-0143542.	PR 04-OCT-1999;	99US-0157117.
PR 14-JUL-1999;	99US-0143624.	PR 05-OCT-1999;	99US-0157753.
PR 15-JUL-1999;	99US-0144005.	PR 06-OCT-1999;	99US-0157865.
PR 16-JUL-1999;	99US-0144085.	PR 07-OCT-1999;	99US-0158029.
PR 16-JUL-1999;	99US-0144086.	PR 08-OCT-1999;	99US-0158232.
PR 19-JUL-1999;	99US-0144325.	PR 12-OCT-1999;	99US-0158369.
PR 19-JUL-1999;	99US-0144331.	PR 13-OCT-1999;	99US-0159293.
PR 19-JUL-1999;	99US-0144332.	PR 13-OCT-1999;	99US-0159294.
PR 19-JUL-1999;	99US-0144333.	PR 13-OCT-1999;	99US-0159295.
PR 19-JUL-1999;	99US-0144334.	PR 14-OCT-1999;	99US-0159329.
PR 19-JUL-1999;	99US-0144335.	PR 14-OCT-1999;	99US-0159330.
PR 20-JUL-1999;	99US-0144352.	PR 14-OCT-1999;	99US-0159331.
PR 20-JUL-1999;	99US-0144632.	PR 14-OCT-1999;	99US-0159637.

PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 37.5%; Score 3; DB 21; Length 164;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
1111
Db 112 syxg 115

RESULT 17
AAR72680
ID AAR72680 standard; peptide; 226 AA.

AC AAR72680;

DT 01-NOV-1995 (first entry)

DE Human plasma glutathione peroxidase.

KW Human glutathione peroxidase; plasma; antibody; immunoassay; kidney;
seelenium deficiency; liver disease; graft rejection; hepatic cancer.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 73 /note="unknown amino acid"

PN WO9506719-A.

PD 09-MAR-1995.

PF 01-SEP-1994; 94WO-FR01031.

PR 03-SEP-1993; 93FR-0010504.

PA (BIOX-) BIOXYTECH SA.
(BIOX-) BIOXYTECH.

PI Chaudiere J, Lemaingue A, Malette P;

XX WPI: 1995-115435/15.

PT New immunogenic Oligo:peptide of human plasma glutathione
peroxidase - and derived antibodies that do not recognise other
PT forms of the enzyme and are useful in immunoassays, e.g. to
PT detect selenium deficiency

PS * Disclosure: Fig 1; 33pp; French.

XX The deduced amino acid sequence of the human plasma glutathione

CC peroxidase as published (J. Biochem., 108, 1990,145-148). Immunogenic
CC peptides (e.g. see AAR72679) were derived from
CC hydropobicity/flexibility profiles of the published sequence. Peptides
CC containing this sequence can be conjugated to carrier proteins and used
CC to produce antibodies. The antibodies can be used to detect plasma
CC glutathione peroxidase in immunoassays e.g. for detection of selenium
CC deficiency or in diagnosis of certain kidney or liver diseases e.g. graft
CC rejection or certain hepatic cancers.

XX Sequence 226 AA;

Query Match 37.5%; Score 3; DB 16; Length 226;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
1111
Db 71 syxg 74

RESULT 18
AAB22828
ID AAB22828 standard; Protein; 226 AA.

AC AAB22828;

DT 10-JAN-2001 (first entry)

DE Human plasma glutathione peroxidase H (pGPXH).

KW Plasma glutathione peroxidase H; pGPXH; human;
recombinant production.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 73 /label= unknown
FT /note="Encoded by TGA"

PN CN1256312-A.

PD 14-JUN-2000.

PF 29-OCT-1998; 98CN-0121973.

PR 29-OCT-1998; 98CN-0121973.

PA (UYFU-) UNITV FUDAN.

XX Yu L, Tu Q, Fu Q;

DR WPI: 2000-533631/49.

DR N-PSDB; AAA90423.

PT New human glutathione peroxidase and its code sequence, preparation and
PT use -

PS Claim 2; Page 20-21; 26pp; Chinese.

CC This sequence represents human plasma glutathione peroxidase H
CC (pGPXH). The invention relates to this novel human glutathione
CC peroxidase, nucleic acid encoding it, and to recombinant production
CC of human pGPXH. The invention also encompasses applications for
CC human pGPXH.

XX Sequence 226 AA;

Query Match 37.5%; Score 3; DB 21; Length 226;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
||||
DB 71 syxg 74

RESULT 19
AAW08430
ID AAW08430 standard; Protein: 230 AA.
XX
AC AAW08430;
XX
DT 10-OCT-1997 (first entry)
XX
DE Rapamycin-dependent activation of signal transduction fusion T6.
XX
KW Rapamycin: responsive transfected cell; recombinant DNA;
KW fusion protein; rapamycin binding domain; RBD; FRBP;
KW FK506-binding protein; FRB domain; genetic engineering; regulation;
KW transcription; signal transduction; apoptosis; inhibition; virus;
KW gene therapy; ss.
XX
OS Synthetic.
XX
FN WO9641865-A1.
XX
PD 27-DEC-1996.
XX
PE 07-JUN-1996; 96WO-US09948.
XX
PR 09-FEB-1996; 96US-0598776.
PR 07-JUN-1995; 95US-0481941.
XX
PA (ARIA-) ARIAD GENE THERAPEUTICS INC.
XX
PI Clackson T, Gilman MZ, Holt DA;
DR WPI: 1997-063453/06.
DR N-PSDB: AAT49061.
XX
PT New rapamycin-responsive transfected cells - contg. recombinant DNA
PT encoding fusion proteins which act as biological switches for
PT regulating biological events
XX
XX Example 8; Page 109; 149pp; English.
XX
PS The present sequence represents the representative construct fusion T6
CC which is involved in the rapamycin-dependent activation of signal
CC transduction. This is involved in the construction of transcription
CC factor fusion proteins, which in turn are involved in the production of
CC animal cells containing at least two recombinant DNAs. One DNA should
CC encode a chimeric protein which is capable of binding to rapamycin, or a
CC rapamycin analogue, and comprises at least one rapamycin binding domain
CC (RBD) and at least one protein domain heterologous to it; the second
CC recombinant DNA should encode a second chimeric protein which is capable
CC of forming a complex with rapamycin, or a rapamycin analogue, and the
CC first chimeric protein and comprises at least one FKBP:rapamycin binding
CC (FRB) domain and at least one domain heterologous to it. Contacting
CC these genetically engineered cells with rapamycin or analogues results
CC in the formation of a complex (between the fusion proteins, and the
CC rapamycin), and initiation of a biological response. The products can be
CC used for regulating biological events such as gene transcription,
CC activation of an intracellular signal transduction pathway leading to
CC e.g. gene expression or apoptotic cell death, gene knock-out, blockade
CC of a gene or inhibition of function of a gene product. They are used
CC particularly for regulated gene therapy and for production of
CC recombinant proteins and viruses.
XX
SQ Sequence 230 AA;

Query Match 37.5%; Score 3; DB 18; Length 230;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 syxg 8
||||
DB 226 syxg 229

RESULT 20
AAB08728
ID AAB08728 standard; Protein: 280 AA.
XX
AC AAB08728;
XX
DT 02-JAN-2001 (first entry)
XX
DE Amino acid sequence of a consensus B7RP1 polypeptide.
XX
KW CRP1; CD28 related protein-1; B7RP1; B7 related protein-1;
KW T-lymphocyte activation; type I transmembrane protein; T cell activation;
KW T cell proliferation; T-cell mediated disorder.
XX
OS Synthetic.
OS Mus sp.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT MISC-difference 1..280
FT FT
XX
PN WO200046240-A2.
XX
PD 10-AUG-2000.
XX
PE 27-JAN-2000; 2000WO-US01871.
XX
PR 03-FEB-1999; 99US-0244448.
PR 08-MAR-1999; 99US-0264527.
XX
PA (AMGE-) AMGEN INC.
XX
PI Yoshinaga SK;
DR WPI: 2000-543476/49.
XX
PT Novel nucleic acids encoding the proteins CRP-1 and B7RP1 are useful
PT in the treatment, prevention and diagnosis of T cell mediated disorders
PT -
XX
PS Disclosure; Page 161-162; 174pp; English.
XX
CC The present sequence represents a consensus B7RP1 (B7 related protein-1)
CC polypeptide. The specification also describes a CRP1 (CD28 related
CC protein-1) polypeptide. The polypeptides are involved in T-lymphocyte
CC activation, and represent a receptor-ligand pair. CRP1 and B7RP1 are
CC predicted to be a type I transmembrane protein. The nucleic acids are
CC useful for regulating T cell activation or proliferation in an animal.
CC The polypeptides are useful for treating, preventing ameliorating or
CC diagnosing a T-cell mediated disorder in an animal. They can also be
CC used to identify test molecules that bind to the polypeptides.
XX
SQ Sequence 280 AA;

Query Match 37.5%; Score 3; DB 21; Length 280;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 syxg 8
||||
DB 265 syxg 268

RESULT 21

AAB08730
ID AAB08730 standard; Protein: 329 AA.
AC AAB08730;
XX
DT 02-JAN-2001 (first entry)
XX
DE Amino acid sequence of a consensus B7RP1 polypeptide.
XX
KW CRP1: CD28 related protein-1; B7RP1: B7 related protein-1;
XX T-lymphocyte activation; type I transmembrane protein; T cell activation;
XX T cell proliferation; T-cell mediated disorder.
XX
OS Synthetic.
OS Mus sp.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT MISC-difference 1..329
FT /note= "Xaa represent unspecified amino acids"
XX
XX WO200046240-A2.
XX
XX 10-AUG-2000.
XX
XX 27-JAN-2000; 2000WO-US01871.
XX
XX 03-FEB-1999; 99US-0244448.
XX 08-MAR-1999; 99US-0264527.
XX
XX (AMGE-) AMGEN INC.
XX
XX Yoshinaga SK;
XX
XX WPI: 2000-543476/49.
XX
XX Novel nucleic acids encoding the proteins CRP-1 and B7RP1 are useful
XX in the treatment, prevention and diagnosis of T cell mediated disorders
XX
XX
XX Disclosure: Page 167-168; 174pp; English.
XX
XX The present sequence represents a consensus B7RP1 (B7 related protein-1)
XX polypeptide. The specification also describes a CRP1 (CD28 related
XX protein-1) polypeptide. The polypeptides are involved in T-lymphocyte
XX activation, and represent a receptor-ligand pair. CRP1 and B7RP1 are
XX predicted to be a type I transmembrane protein. The nucleic acids are
XX useful for regulating T cell activation or proliferation in an animal.
XX The polypeptides are useful for treating, preventing ameliorating or
XX diagnosing a T-cell mediated disorder in an animal. They can also be
XX used to identify test molecules that bind to the polypeptides.
XX
SQ Sequence 329 AA:

Query Match 37.5%; Score 3; DB 21; Length 329;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 syxg 8
|
|
|
|
Db 312 syxg 315

RESULT 22
ID AAR44893 standard; Protein: 533 AA.
XX
XX AAR44893;
XX
XX 22-JUN-1994 (first entry)
XX
XX Diphtheria toxin (delta-147-148; H21X) mutant.
XX
XX

XX
XX DT; protein exotoxin; NAD-dependent ADP-ribosyltransferase; vaccine;
KW diphtheria toxoid; deletion mutant; mutin; variant; double mutant;
KW reversion mutation; site-directed mutagenesis.
XX
XX Corynebacterium diphtheriae.
OS
XX
XX
FH Key Location/Qualifiers
FT Protein 1..533
FT /note= "diphtheria toxin mutant; Val(147) and
FT Glu(148) have been deleted and His(21)
FT is substid. by any other amino acid or
FT is absent"
FT
FT MISC-difference 21
FT /note= "any amino acid other than wild-type His
FT or absent"
XX
XX WO9325210-A.
XX
XX 23-DEC-1993;
XX
XX 17-MAY-1993; 93WO-US04606.
XX
XX 18-JUN-1992; 92US-0901712.
XX
XX (HARD) HARVARD COLLEGE.
XX
XX Collier RJ, Killien K, Mekalanos J;
XX WPI: 1994-007178/01.
XX
XX N-PSDB: AAO54341.
XX
XX New DNA encoding diphtheria toxin deletion mutants - with no
XX toxicity and low risk of reversion, and derived toxoids and
XX transformed cells, useful in vaccines
XX
XX Claim 11; : 42pp; English.
XX
XX Oligonucleotide-directed mutagenesis of the wild-type diphtheria
XX gene results in deletion of the codons for Val-147 and active site
XX residue Glu-148. The resulting mutin is not toxic, making it
XX useful in diphtheria vaccines. The risk of reversion to toxicity
XX is much lower for the 147-148 double mutant than for the prior art
XX 148 single mutant, while its immunogenicity is not impaired. The
XX 147-148 mutin opt. has other amino acid residues substid. or
XX deleted, e.g. wild-type His(21) in fragment A. The specification
XX includes the wild-type DT amino acid sequence (see AAR44888) but does
XX not include any mutant sequences; the wild-type sequence was modified
XX according to the description in the claims to give AAR44893.
XX
SQ Sequence 533 AA:

Query Match 37.5%; Score 3; DB 15; Length 533;
Best Local Similarity 100.0%; Pred. No. 2.5e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 syxg 8
|
|
|
|
Db 19 syxg 22

RESULT 23
ID AAW17581 standard; protein: 684 AA.
XX
XX AAW17581;
XX
XX 01-JUL-1997 (first entry)
XX
XX Thermoaerobacter CGTase variant beta-cyclodextrin #1.
XX
XX Cyclomaltoextrin glucanotransferase; CGTase; enzyme; Bacillus circulans;
XX
XX

PA (NOVO) NOVO-NORDISK AS.
XI Andersen C, Dijkhuizen L, Dijkstra BW, Von Der Osten C;
XX WPI; 1996-485774/48.
XX
PT New variants of cyclo:malto:dextrin glucanotransferase (CGTase) -
PT have altered substrate binding, useful for prodn. of cyclodextrin(s)
PT or linear oligosaccharide(s), opt. formed in situ in e.g. baked
PT goods
XX
PS Claim 31; Page -: 161pp; English.
XX
CC AAM17580-W17591, and AAM17606-W17652 represent mutant versions of the
CC cyclomalto-dextrin glucanotransferase (CGTase) of Thermoaerobacter
CC thermophilurigenes sp. ATCC 53627. (see AAM06772 for wild type
CC sequence). CGTase catalyses the conversion of starch and similar
CC substrates into cyclomalto-dextrins (also known as cyclodextrins) via an
CC intramolecular transglycosylation reaction. Cyclodextrins (CD) are cyclic
CC glucose oligomers with a hydrophobic internal cavity that form inclusion
CC complexes with many small hydrophobic molecules. These CGTase mutants
CC have a modified substrate binding and/or product selectivity, compared
CC to this sequence. The mutants are created using primer mutagenesis to
CC modify the gene encoding this sequence. These sequences have greater
CC product selectivity and/or reduced product inhibition (better yields)
CC than wild-type CGTase. These mutant sequences are used to manufacture the
CC 6 main CD types (comprising 6-11 glucose units), or linear 2-12 glucose
CC oligosaccharides, optionally in situ, e.g. during production of baked
CC goods, to stabilise chemicals during their manufacture and in
CC detergents. CD are known for their usefulness in foods, e.g. as a
CC bread-improving agent, to encapsulate/stabilise/solubilise vitamins,
CC dyes, pharmaceuticals, pesticides or fungicides, to bind/remove
CC lipophilic compounds such as cholesterol (e.g. in egg yolk or butter), in
CC plastic laminates, films etc. and to make biodegradable plastics.
XX
SQ Sequence 684 AA:

Query Match 37.5%; Score 3; DB 17; Length 684;
Best Local Similarity 100.0%; Pred. NO. 2.9e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
|||
Db 97 syxg 100

RESULT 24
AAM17585
ID AAM17585 standard; protein; 684 AA.
XX
AC AAM17585;
XX
DT 01-JUL-1997 (first entry)
XX
DE Thermoaerobacter CGTase variant beta-cyclodextrin #2.
XX
KW Cyclomalto-dextrin glucanotransferase; CGTase; enzyme; Bacillus circulans;
KW thermoaerobacter; starch; cyclomalto-dextrin; cyclodextrin; pesticide;
KW transglycosylation; cyclic glucose oligomer; glucose oligosaccharide;
KW cholesterol removal; food stabiliser; vitamin stabiliser; fungicide;
KW plastic laminate; biodegradable plastic; muten.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 21 /label= V21X
FT /note= "X= Phe, Tyr"
FT Modified-site 47 /label= K47X
FT /note= "X= Gln, Leu, Ala, His, Arg"
FT Modified-site 87

FT /label= L87H
FT Modified-site 88 /label= P88X
FT /note= "X= Pro, Asn, Lys, His"
FT Modified-site 89 /label= D89X
FT /note= "X= Asp, Gly, Ala, Lys, Arg, Pro, Glu, or absent"
FT Modified-site 90 /label= S90X
FT /note= "X= Gly, Ala"
FT Modified-site 91 /label= T91X
FT /note= "X= Ala, Val, Ser, Thr"
FT Misc-difference 92 /label= P91X
FT /note= "X= Ala, Val, Gly, Phe, Tyr, or absent"
FT Modified-site 94 /label= G93X
FT /note= "X= Gly, His, Thr or absent"
FT Modified-site 95 /label= S94X
FT /note= "X= Gln, Lys, Arg, Trp, Phe, Ser, or absent"
FT Modified-site 99 /label= H98X
FT /note= "X= Gly, Ala"
FT Modified-site 102 /label= W101X
FT /note= "X= Gly, Ala"
FT Modified-site 136 /label= D135L
FT Modified-site 141 /label= H140X
FT /note= "X= Ala, Arg, Asn"
FT Modified-site 146 /label= S145X
FT /note= "X= Ala, Glu, Trp, Leu"
FT Misc-difference 147 /note= "X= Ala, Glu, Trp, Leu"
FT Modified-site 148 /label= E146X
FT /note= "X= Pro, Ala, Phe, Gln, Ser, Trp, Ile, Arg, Glu, Lys, Asp, Asn, or absent"
FT Modified-site 149 /label= T147X
FT /note= "X= Ala, Leu, Ile, Phe, Trp, Gly, Tyr, Arg, Asp, Thr, or absent"
FT Modified-site 150 /label= D148X
FT /note= "X= Gly, Asn, Ala"
FT Modified-site 151 /label= P149W
FT Modified-site 152 /label= T150A
FT Modified-site 169 /label= Y167X
FT /note= "X= Ala, Phe"
FT Modified-site 170 /label= T168S
FT Modified-site 180 /label= Y178N
FT Modified-site 181 /label= G179X
FT /note= "X= Ser, Asn, Asp"
FT Modified-site 182 /label= G180X
FT /note= "X= Ser, Asn, Asp"
FT Modified-site 185 /label= F183X
FT /note= "X= Trp, Tyr, Ala"
FT Modified-site 187 /label= S185X
FT /note= "X= Pro, His, Arg, Glu, Asp"
FT Modified-site 194

FT	Modified-site	/label= R192K
FT	Modified-site	/label= N193X
FT	Modified-site	/note= "X: Gly, Ala"
FT	Modified-site	197
FT	Modified-site	/label= F195X
FT	Modified-site	/note= "X: Leu, Ile, Trp, Phe"
FT	Modified-site	198
FT	Modified-site	/label= D196X
FT	Modified-site	/note= "X: Ala, Gly, Asn, Ser"
FT	Modified-site	199
FT	Modified-site	/label= L197X
FT	Modified-site	/note= "X: Asp, Glu"
FT	Modified-site	234
FT	Modified-site	/label= K232X
FT	Modified-site	/note= "X: Glu, Leu"
FT	Modified-site	235
FT	Modified-site	/label= H233X
FT	Modified-site	/note= "X: Glu, Asn, Ile"
FT	Modified-site	261
FT	Modified-site	/label= Y259X
FT	Modified-site	/note= "X: Phe, Trp, Ala"
FT	Modified-site	266
FT	Modified-site	/label= E264Q
FT	Modified-site	328
FT	Modified-site	/label= N326X
FT	Modified-site	/note= "X: Glu, Phe, Leu"
FT	Modified-site	371
FT	Modified-site	/label= G370X
FT	Modified-site	/note= "X: Thr, Asn"
FT	Modified-site	372
FT	Modified-site	/label= D371X
FT	Modified-site	/note= "X: Ala, Ser, Asn, Gly"
FT	Modified-site	374
FT	Modified-site	/label= Y373X
FT	Modified-site	/note= "X: Asp, Glu, Tyr"
FT	Modified-site	376
FT	Modified-site	/label= R375X
FT	Modified-site	/note= "X: Ala, Pro, Gly, Lys"
FT	Modified-site	598
FT	Modified-site	/label= L600X
FT	Modified-site	/note= "X: unspecified amino acid"
XX		W09633267-A1.
XX		24-OCT-1996.
PD		
XX		22-APR-1996; 96WO-DK00179.
PF		
XX		16-NOV-1995; 95DK-0001281.
PR		21-APR-1995; 95DK-0000477.
PR		17-OCT-1995; 95DK-0001173.
XX		(NOVO) NOVO-NORDISK AS.
PA		
XX		
PI	Andersen C, Dijkhuizen L, Dijkstra BW, Von Der Osten C;	
XX		
DR	WPI: 1996-485774/48.	
XX		
XX	New variants of cyclo:maltodextrin glucanotransferase (CGTase) -	
PT	have altered substrate binding, useful for prodn. of cyclodextrin(s)	
PT	or linear oligosaccharide(s), opt. formed in situ in e.g. baked	
PT	goods	
XX		
PS	Claim 34; Page -: 161pp; English.	
XX		
CC	AAW17580-W17591, and AAW17606-W17652 represent mutant versions of the	
CC	cyclomalto-dextrin glucanotransferase (CGTase) of Thermoaerobacter	
CC	thermosulphurigenes sp. ATCC 53627. (see AAW06772 for wild type	
CC	sequence). CGTase catalyses the conversion of starch and similar	
CC	substrates into cyclomalto-dextrins (also known as cyclodextrins) via an	
CC	intramolecular transglycosylation reaction. Cyclodextrins (CD) are cyclia	
CC	glucose oligomers with a hydrophobic internal cavity that form inclusion	

Query Match	Best Local Similarity	37.5%;	Score 3;	DB 17;	Length 684;
Matches 4;	Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	5 syxg 8 				
Db	97 syxg 100				
RESULT 25					
AAW17589					
ID	AAW17589 standard; protein; 684 AA.				
XX					
AC	AAW17589;				
XX					
DF	01-JUL-1997 (first entry)				
XX					
DE	Thermoanaerobacter CGTase variant beta-cyclodextrin #3.				
XX					
KW	Cyclomaltopectrin glucanotransferase; CGTase; enzyme; Bacillus circulans;				
KW	thermoanaerobacter; starch; cyclomaltopectrin; cyclodextrin; pesticide;				
KW	transglycosylation; cyclic glucose oligomer; glucose oligosaccharide;				
KW	cholesterol removal; food stabiliser; vitamin stabiliser; fungicide;				
KW	plastic laminate; biodegradable plastic; muten.				
OS	Synthetic.				
XX					
FH	Key				
FT	Modified-site	21	Location/Qualifiers		
FT		/label= V21X			
FT		/note="X= Phe, Tyr"			
FT	Modified-site	47			
FT		/label= K47X			
FT		/note="X= Gln, Leu, Ala, His, Arg"			
FT	Modified-site	87			
FT		/label= L87X			
FT		/note="X= Ile, His"			
FT	Modified-site	88			
FT		/label= P88X			
FT		/note="X= Asn, Lys, His"			
FT	Modified-site	89			
FT		/label= D89X			
FT		/note="X= Gly, Ala, Tyr, Lys, Arg, Pro, Glu, or absent"			
FT	Modified-site	90			
FT		/label= S90X			
FT		/note="X= Gly, Ala"			
FT	Modified-site	91			
FT		/label= T91X			
FT		/note="X= Ala, Val, Ser, Gly"			
FT	Misc-difference	92			
FT		/label= F91aX			
FT		/note="X= Ala, Val, Gly, Tyr, or absent"			
FT	Modified-site	93			
FT		/label= G92V			
FT	Modified-site	94			

FT /label= G93X
FT /note= "X= Asn, His, Thr or absent"
FT Modified-site
FT /label= S94X
FT /note= "X= Gln, Lys, Arg, Trp, Phe, Asn, or absent"
FT Modified-site
FT /label= H98X
FT /note= "X= Gly, Ala"
FT Modified-site
FT /label= W101X
FT /note= "X= Gly, Ala"
FT Modified-site
FT /label= D135L
FT /label= H140X
FT /note= "X= Ala, Arg, Asn"
FT Modified-site
FT /label= S145X
FT /note= "X= Ala, Glu, Trp, Leu"
FT Misc-difference
FT /label= 147
FT /note= "insertion of Pro, Ala, Ile, Gln, or Ser"
FT Modified-site
FT /label= E146X
FT /note= "X= Pro, Ala, Phe, Gln, Ser, Trp, Ile, Arg, Lys, Asp, Asn, or absent"
FT Modified-site
FT /label= T147X
FT /note= "X= Ala, Leu, Ile, Phe, Trp, Gly, Tyr, Arg, Asp, or absent"
FT Modified-site
FT /label= D148X
FT /note= "X= Gly, Asn, Ala, Gln"
FT Modified-site
FT /label= P149W
FT /label= T150X
FT /label= "X= Ala, Ser"
FT Modified-site
FT /label= Y167X
FT /note= "X= Ala, Phe"
FT Modified-site
FT /label= T168S
FT /label= Y178N
FT /label= G179X
FT /label= "X= Ser, Asn, Asp"
FT Modified-site
FT /label= G180X
FT /note= "X= Ser, Asn, Asp"
FT Modified-site
FT /label= F183X
FT /note= "X= Trp, Tyr, Ala"
FT Modified-site
FT /label= S185X
FT /note= "X= Pro, His, Arg, Glu, Asp"
FT Modified-site
FT /label= R192K
FT /label= N193X
FT /note= "X= Gly, Ala"
FT Modified-site
FT /label= F195X
FT /note= "X= Leu, Ile, Trp, Tyr"
FT Modified-site
FT /label= D196X
FT /note= "X= Ala, Gly, Asn, Ser"
FT Modified-site
FT /label= L197X
FT /note= "X= Asp, Glu"
FT Modified-site
FT /label= K232X
FT /note= "X= Gln, Leu"

FT Modified-site 235
FT /label= H233X
FT /note= "X= Gln, Asn, Ile"
FT Modified-site 261
FT /label= Y259X
FT /note= "X= Phe, Trp, Ala"
FT Modified-site 266
FT /label= E264Q
FT Modified-site 328
FT /label= N326X
FT /note= "X= Gln, Phe, Leu"
FT Modified-site 371
FT /label= G370X
FT /note= "X= Thr, Asn"
FT Modified-site 372
FT /label= D371X
FT /note= "X= Ala, Ser, Asn, Gly"
FT Modified-site 374
FT /label= Y373X
FT /note= "X= Asp, Glu"
FT Modified-site 376
FT /label= R375X
FT /note= "X= Ala, Pro, Gly, Lys"
FT Modified-site 598
FT /label= L600X
FT /note= "X= unspecified amino acid"
XX W09633267-A1.
PN
XX
XX 24-OCT-1996.
PD
XX
XX 22-APR-1996; 96WO-DK00179.
PE
XX
XX 16-NOV-1995; 95DK-0001281.
PR 21-APR-1995; 95DK-0000477.
PR 17-OCT-1995; 95DK-0001173.
XX
XX (NOVO) NOVO-NORDISK AS.
XX
XX Andersen C, Dijkhuizen L, Dijkstra BW, Von Der Osten C;
PI
XX
XX WPI; 1996-485774/48.
DR
XX
XX New variants of cyclo:malto:dextrin glucanotransferase (CGTase) -
PT have altered substrate binding, useful for prodn. of cyclodextrin(s)
PT or linear oligosaccharide(s), opt. formed in situ in e.g. baked
PT goods
XX
XX
PS Claim 40; Page -: 161pp; English.
XX
XX AAW17580-W17591, and AAW17606-W17652 represent mutant versions of the
CC cyclomaltodextrin glucanotransferase (CGTase) of Thermooanoerobacter
CC thermophilus sp. ATCC 53627. (see AAW06772 for wild type
CC sequence). CGTase catalyses the conversion of starch and similar
CC substrates into cyclomaltodextrins (also known as cyclodextrins) via an
CC intramolecular transglycosylation reaction. Cyclodextrins (CD) are cyclic
CC glucose oligomers with a hydrophobic internal cavity that form inclusion
CC complexes with many small hydrophobic molecules. These CGTase mutants
CC have a modified substrate binding and/or product selectivity, compared
CC to this sequence. The mutants are created using primer mutagenesis to
CC modify the gene encoding this sequence. These sequences have greater
CC product selectivity and/or reduced product inhibition (better yields)
CC than wild-type CGTase. These mutant sequences are used to manufacture the
CC 6 main CD types (comprising 6-11 glucose units), or linear 2-12 glucose
CC oligosaccharides, optionally in situ, e.g. during production of baked
CC goods, to stabilise chemicals during their manufacture and in
CC detergents. CD are known for their usefulness in foods, e.g. as a
CC bread-improving agent, to encapsulate/stabilise/solubilise vitamins,
CC dyes, pharmaceuticals, pesticides or fungicides, to bind/remove
CC lipophilic compounds such as cholesterol (e.g. in egg yolk or butter), in
CC plastic laminates, films etc. and to make biodegradable plastics.
XX
XX Sequence 684 AA;

Query Match 37.5%; Score 3; DB 17; Length 684;
Best Local Similarity 100.0%; Pred. No. 2.9e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
||||
DB 97 syxg 100

Search completed: January 14, 2002, 07:56:28
Job time: 723 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: January 14, 2002, 07:51:15 ;

Search time 41.59 Seconds
(without alignments)
4.329 Million cell updates/sec

Title: 09-185908-1A

Perfect score: 8

Sequence: 1 wxxsxyxg 8

Scoring table: OLIGO

Gapco 60.0 , Gapext 60.0

Searched: 212252 seqs, 22503292 residues

Word size : 0

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

Databasc :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCRTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3	37.5	104	1	US-08-488-113B-153
2	3	37.5	104	1	US-08-477-484B-153
3	3	37.5	104	1	US-08-107-668D-17
4	3	37.5	104	1	US-08-472-788A-17
5	3	37.5	104	2	US-08-477-531B-17
6	3	37.5	104	2	US-08-646-360-153
7	3	37.5	104	2	US-08-082-842A-17
8	3	37.5	104	4	US-08-839-765-153
9	3	37.5	104	4	US-09-136-389-153
10	3	37.5	226	2	US-08-428-188-1
11	3	37.5	526	4	US-08-895-590-5
12	2	25.0	3	4	US-09-461-697-405
13	2	25.0	4	1	US-07-895-300A-8
14	2	25.0	4	1	US-08-079-445-3
15	2	25.0	4	1	US-07-840-077A-7
16	2	25.0	4	1	US-08-332-071B-16
17	2	25.0	4	1	US-08-176-938-20
18	2	25.0	4	1	US-08-487-006-150
19	2	25.0	4	1	US-08-487-006-151
20	2	25.0	4	1	US-08-487-006-152
21	2	25.0	4	1	US-08-487-006-153
22	2	25.0	4	1	US-08-170-360-17
23	2	25.0	4	1	US-08-454-950-7
24	2	25.0	4	1	US-08-434-761-3
25	2	25.0	4	1	US-08-338-890B-1
26	2	25.0	4	1	US-08-454-949-7
27	2	25.0	4	1	US-08-458-367-8

28	2	25.0	4	2	US-08-441-871-1	Sequence 1, Appl1
29	2	25.0	4	2	US-08-441-871-55	Sequence 55, Appl1
30	2	25.0	4	2	US-08-441-871-65	Sequence 65, Appl1
31	2	25.0	4	2	US-08-685-589A-6	Sequence 6, Appl1
32	2	25.0	4	2	US-08-488-659A-150	Sequence 150, App
33	2	25.0	4	2	US-08-488-659A-151	Sequence 151, App
34	2	25.0	4	2	US-08-488-659A-152	Sequence 152, App
35	2	25.0	4	2	US-08-488-659A-153	Sequence 153, App
36	2	25.0	4	2	US-08-671-487A-8	Sequence 8, Appl1
37	2	25.0	4	2	US-08-645-193B-27	Sequence 27, Appl1
38	2	25.0	4	2	US-08-340-208B-1	Sequence 1, Appl1
39	2	25.0	4	2	US-08-722-806A-7	Sequence 7, Appl1
40	2	25.0	4	2	US-08-651-179B-2	Sequence 2, Appl1
41	2	25.0	4	2	US-08-651-179B-3	Sequence 3, Appl1
42	2	25.0	4	3	US-08-997-263-1	Sequence 1, Appl1
43	2	25.0	4	3	US-09-222-373-49	Sequence 49, Appl1
44	2	25.0	4	3	US-09-222-373-50	Sequence 50, Appl1
45	2	25.0	4	3	US-09-315-861-5	Sequence 5, Appl1
46	2	25.0	4	3	US-09-112-656-14	Sequence 14, Appl1
47	2	25.0	4	3	US-08-981-122-25	Sequence 25, Appl1
48	2	25.0	4	4	US-08-435-568A-19	Sequence 19, Appl1
49	2	25.0	4	4	US-09-193-365-14	Sequence 14, Appl1
50	2	25.0	4	4	US-09-202-621B-7	Sequence 7, Appl1

ALIGNMENTS

RESULT 1
US-08-488-113B-153
Sequence 153, Application US/08488113B
Patent No. 5744580
GENERAL INFORMATION:
APPLICANT: Better, Marc D.
APPLICANT: Carroll, Stephen F.
APPLICANT: Studnika, Gary M.
TITLE OF INVENTION: Immunotoxins Comprising Ribosome-Inactivating
NUMBER OF INVENTION: Proteins
NUMBER OF SEQUENCES: 169
CORRESPONDENCE ADDRESSES:
ADDRESSEE: McAndrews, Held & Malloy, Ltd.
STREET: 500 West Madison Street, 34th floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60661
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,113B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/425,336
FILING DATE: 18-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/064,691
FILING DATE: 12-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/988,430
FILING DATE: 09-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/901,707
FILING DATE: 19-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/787,567
FILING DATE: 04-NOV-1991
ATTORNEY/AGENT INFORMATION:
NAME: McNicholas, Janet M.
REGISTRATION NUMBER: 32,918

REFERENCE/DOCKET NUMBER: 110220US07/200-70.P3.C2A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/707-8889
TELEFAX: 312/707-9155
TELEX: 650 388-1248
INFORMATION FOR SEQ ID NO: 153:
SEQUENCE CHARACTERISTICS:
LENGTH: 104 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-488-113B-153

Query Match 37.5%; Score 3; DB 1; Length 104;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
||||
Db 85 SYXG 88

RESULT 2
US-08-477-484B-153
Sequence 153, Application US/08477484B
Patent No. 5756699
GENERAL INFORMATION:
APPLICANT: Better, Marc D.
APPLICANT: Carroll, Stephen F.
APPLICANT: Studnika, Gary M.
TITLE OF INVENTION: Immunotoxins Comprising Ribosome-Inactivating
NUMBER OF SEQUENCES: 169
CORRESPONDENCE ADDRESSES:
ADDRESSEE: McAndrews, Held & Malloy, Ltd.
STREET: 500 West Madison Street, 34th floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60661
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,484B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/425,336
FILING DATE: 18-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/064,691
FILING DATE: 12-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/988,430
FILING DATE: 09-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/901,707
FILING DATE: 19-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/787,567
FILING DATE: 04-NOV-1991
ATTORNEY/AGENT INFORMATION:
NAME: McNicholas, Janet M.
REGISTRATION NUMBER: 32,918
REFERENCE/DOCKET NUMBER: 110220US07/200-70.P3.C2A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/707-8889
TELEFAX: 312/707-9155

TELEX: 650 388-1248
INFORMATION FOR SEQ ID NO: 153:
SEQUENCE CHARACTERISTICS:
LENGTH: 104 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-477-484B-153

Query Match 37.5%; Score 3; DB 1; Length 104;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
||||
Db 85 SYXG 88

RESULT 3
US-08-107-669D-17
Sequence 17, Application US/08107669D
Patent No. 5766886
GENERAL INFORMATION:
APPLICANT: Studnika, Gary M.
TITLE OF INVENTION: Modified Antibody Variable Domains (as amended)
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein and Fox P.L.L.C.
STREET: 1100 New York Ave., N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/107,669D
FILING DATE: 13-AUG-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/10906
FILING DATE: 14-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/808,464
FILING DATE: 13-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Michele A. Cimbalà
REGISTRATION NUMBER: 33,851
REFERENCE/DOCKET NUMBER: 0610.1000001/MAC
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202/371-2600
TELEFAX: 202/371-2540
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 104 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-107-669D-17

Query Match 37.5%; Score 3; DB 1; Length 104;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
||||
Db 85 SYXG 88

RESULT 4
US-08-472-788A-17
Sequence 17, Application US/08472788A
Patent No. 5770196
GENERAL INFORMATION:
APPLICANT: Studnicka, Gary M.
TITLE OF INVENTION: Modified Antibody Variable Domains
NUMBER OF SEQUENCES: 89
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein and Fox P.L.L.C.
STREET: 1100 New York Ave., N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/472,788A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/082,842
FILING DATE: 23-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/10906
FILING DATE: 14-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/808,464
FILING DATE: 13-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Cimbalia, Michele A.
REGISTRATION NUMBER: 33,851
REFERENCE/DOCKET NUMBER: 0610.1000003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202/371-2600
TELEFAX: 202/371-2540
TELEX:
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 104 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-472-788A-17
Query Match 37.5%; Score 3; DB 1; Length 104;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 syxg 8
|||
Db 85 syxg 88

STATE: D.C.
COUNTRY: United States of America
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,531B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/107,669
FILING DATE: 13-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/10906
FILING DATE: 14-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/808,464
FILING DATE: 13-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Michele A. Cimbalia
REGISTRATION NUMBER: 33,851
REFERENCE/DOCKET NUMBER: 0610.1000004/MAC
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202/371-2600
TELEFAX: 202/371-2540
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 104 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-477-531B-17
Query Match 37.5%; Score 3; DB 2; Length 104;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 syxg 8
|||
Db 85 syxg 88
RESULT 6
US-08-646-360-153
Sequence 153, Application US/08646360
Patent No. 5837491
GENERAL INFORMATION:
APPLICANT: Better, Marc D.
APPLICANT: Carroll, Stephen F.
APPLICANT: Studnicka, Gary M.
TITLE OF INVENTION: Immunotoxins Comprising Ribosome-Inactivating
NUMBER OF SEQUENCES: 173
CORRESPONDENCE ADDRESS:
ADDRESSEE: McAndrews, Held & Malloy, Ltd.
STREET: 500 West Madison Street, 34th floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60661
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/646,360
FILING DATE: 13-MAY-1996
CLASSIFICATION: 530

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/05348
FILING DATE: 12-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/064,691
FILING DATE: 12-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/988,430
FILING DATE: 09-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/901,707
FILING DATE: 19-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/787,567
FILING DATE: 04-NOV-1991
ATTORNEY/AGENT INFORMATION:
NAME: McNicholas, Janet M.
REGISTRATION NUMBER: 32,918
REFERENCE/DOCKET NUMBER: 200-70.P4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/707-8889
TELEFAX: 312/707-9135
TELEX: 650 388-1248
INFORMATION FOR SEQ ID NO: 153:
SEQUENCE CHARACTERISTICS:
LENGTH: 104 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-646-360-153

Query Match 37.5%; Score 3; DB 2; Length 104;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 SYXG 8
||||
DB 85 SYXG 88

RESULT 7
US-08-082-842A-17
Sequence 17, Application US/08082842A
Patent No. 5869619
GENERAL INFORMATION:
APPLICANT: Studnicka, Gary M.
TITLE OF INVENTION: Modified Antibody Variable Domains
NUMBER OF SEQUENCES: 89
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein and Fox P.L.L.C.
STREET: 1100 New York Ave., N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,842A
FILING DATE: 23-JUN-1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/10906
FILING DATE: 14-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/808,464
FILING DATE: 13-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: McNicholas, Janet M.
REGISTRATION NUMBER: 11022US09/200-70.P3.C3
REFERENCE/DOCKET NUMBER: 11022US09/200-70.P3.C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/707-8889

NAME: Cimbala, Michele A.
REGISTRATION NUMBER: 33,851
REFERENCE/DOCKET NUMBER: 0610.1000002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202/371-2600
TELEFAX: 202/371-2540
TELEX:
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 104 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-842A-17

Query Match 37.5%; Score 3; DB 2; Length 104;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 SYXG 8
||||
DB 85 SYXG 88

RESULT 8
US-08-839-765-153
Sequence 153, Application US/08839765
Patent No. 614631
GENERAL INFORMATION:
APPLICANT: Better, Marc D.
APPLICANT: Carroll, Stephen F.
APPLICANT: Studnicka, Gary M.
TITLE OF INVENTION: Immunotoxins Comprising Ribosome-Inactivating
NUMBER OF SEQUENCES: 169
CORRESPONDENCE ADDRESS:
ADDRESSEE: McAndrews, Held & Malloy, Ltd.
STREET: 500 West Madison Street, 34th floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60661
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,765
FILING DATE: 15-APR-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/425,336
FILING DATE: 18-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/064,691
FILING DATE: 12-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/988,430
FILING DATE: 09-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/901,707
FILING DATE: 19-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/787,567
FILING DATE: 04-NOV-1991
ATTORNEY/AGENT INFORMATION:
NAME: McNicholas, Janet M.
REGISTRATION NUMBER: 32,918
REFERENCE/DOCKET NUMBER: 11022US09/200-70.P3.C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/707-8889

TELEFAX: 312/707-9155
TELEX: 650 388-1248
INFORMATION FOR SEQ ID NO: 153:
SEQUENCE CHARACTERISTICS:
LENGTH: 104 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-839-765-153

Query Match 37.5%; Score 3; DB 4; Length 104;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
||||
DB 85 SYXG 88

RESULT 9
US-09-136-389-153
Sequence 153, Application US/09136389
Patent No. 6146850
GENERAL INFORMATION:
APPLICANT: Better, Marc D.
APPLICANT: Carroll, Stephen F.
APPLICANT: Studilka, Gary M.
TITLE OF INVENTION: Immunotoxins Comprising Ribosome-Inactivating
NUMBER OF INVENTION: 173
NUMBER OF SEQUENCES: 173
CORRESPONDENCE ADDRESS:
ADDRESSEE: McAndrews, Held & Malloy, Ltd.
STREET: 500 West Madison Street, 34th floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60661
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,389
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/646,360
FILING DATE: 13-MAY-1996
APPLICATION NUMBER: PCT/US94/05348
FILING DATE: 12-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/064,691
FILING DATE: 12-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/988,430
FILING DATE: 09-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/901,707
FILING DATE: 19-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/787,567
FILING DATE: 04-NOV-1991
ATTORNEY/AGENT INFORMATION:
NAME: McNicholas, Janet M.
REGISTRATION NUMBER: 32,918
REFERENCE/DOCKET NUMBER: 200-70.P4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/707-8889
TELEFAX: 312/707-9155
TELEX: 650 388-1248

INFORMATION FOR SEQ ID NO: 153:
SEQUENCE CHARACTERISTICS:
LENGTH: 104 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-136-389-153

Query Match 37.5%; Score 3; DB 4; Length 104;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
||||
DB 85 SYXG 88

RESULT 10
US-08-428-188-1
Sequence 1, Application US/08428188
Patent No. 5861262
GENERAL INFORMATION:
APPLICANT: Chaudhrie, Jean
APPLICANT: Lemaigne, Arnaud
APPLICANT: Malette, Patricia
TITLE OF INVENTION: Method for the Specific Immunoassay of
TITLE OF INVENTION: Human Plasma Glutathione Peroxidase, Kit for Its
TITLE OF INVENTION: Implementation, Oligopeptides and Antibodies Specific for
NUMBER OF INVENTION: 1
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jacobson, Price, Holman & Stern
STREET: 400 Seventh St. N.W.
CITY: Washington D.C.
COUNTRY: U.S.A.
ZIP: 20007
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/428,188
FILING DATE: 02-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 93 10504
FILING DATE: 03-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Holman, John C.
REGISTRATION NUMBER: 22,769
REFERENCE/DOCKET NUMBER: 7696/P58648NA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 638-6666
TELEFAX: (202) 393-5350
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 226 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
FEATURE:
NAME/KEY: Modified-site
LOCATION: 73
OTHER INFORMATION: /note= "Amino acid at position 73
OTHER INFORMATION: is selenocysteine"
US-08-428-188-1

Query Match 37.5%; Score 3; DB 2; Length 226;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
||||
Db 71 SYXG 74

RESULT 11

US-08-895-590-5
; Sequence 5, Application US/08895590
; Patent No. 6207410
; GENERAL INFORMATION:
; APPLICANT: Hall, Linda M.
; APPLICANT: Ren, Dejian
; APPLICANT: Zheng, Wei
; APPLICANT: Dubald, Manuel Marcel Paul
; TITLE OF INVENTION: Genes Encoding an Insect Calcium Channel
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, LLP
; STREET: 699 Prince Street
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22314-3187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/895,590
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,888
; FILING DATE: 19-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm M.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 022650-263
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 526 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-895-590-5

Query Match 37.5%; Score 3; DB 4; Length 526;
Best Local Similarity 100.0%; Pred. No. 6.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
||||
Db 72 SYXG 75

RESULT 12

US-09-461-697-405
; Sequence 405, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.

; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Purnam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 405
; LENGTH: 3
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-405

Query Match 25.0%; Score 2; DB 4; Length 3;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
||
Db 2 WK 3

RESULT 13

US-07-895-300A-8
; Sequence 8, Application US/07895300A
; Patent No. 5279823
; GENERAL INFORMATION:
; APPLICANT: Frenz, John
; APPLICANT: Shire, Steven J.
; APPLICANT: Sliwowski, Mary B.
; TITLE OF INVENTION: PURIFIED FORMS OF DNase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 KB floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/895,300A
; FILING DATE: 19920608
; CLASSIFICATION: 424
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hensley, Max D.
; REGISTRATION NUMBER: 27,043
; REFERENCE/DOCKET NUMBER: 747
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
US-07-895-300A-8

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sy 6
||
Db 2 sy 3

RESULT 14
US-08-079-445-3
; Sequence 3, Application US/08079445
; Patent No. 5440016
; GENERAL INFORMATION:
; APPLICANT: Blondelle, Sylvie E.
; APPLICANT: Pinilla, Clemencia
; APPLICANT: Eichler, Julia
; APPLICANT: Houghten, Richard A.
; TITLE OF INVENTION: PEPTIDES HAVING ANTI-MICROBIAL,
; TITLE OF INVENTION: HEOLYTIC, ANTI-MELITTIN OR ANTI-ENZYMATIC ACTIVITY
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States of America
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/079,445
; FILING DATE: 18-JUN-1993
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-HP 9648
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: /label= Xaa
; OTHER INFORMATION: /note= "Xaa-KFmoc"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 4
; OTHER INFORMATION: /label= Xaa
; OTHER INFORMATION: /note= "Xaa-any amino acid"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 4
; OTHER INFORMATION: /note= "C-terminal amino acid is
; OTHER INFORMATION: amidated"
US-08-079-445-3

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wk 2
||
Db 2 wk 3

RESULT 15
US-07-840-077A-7
; Sequence 7, Application US/07840077A
; Patent No. 5443816
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT: Zamora, Paul O.
; APPLICANT: Rhodes, Buck A.
; TITLE OF INVENTION: Peptide-Metal Ion
; TITLE OF INVENTION: Pharmaceutical Preparation and Method
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhomed Incorporated
; STREET: 4261 Balloun Park
; CITY: Albuquerque
; STATE: NM
; COUNTRY: U.S.A.
; ZIP: 87109-5802
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb
; MEDIUM TYPE: Storage
; COMPUTER: IBM PC/XT/AT, IBM PS/2 or
; OPERATING SYSTEM: PC-DOS or MS-DOS
; SOFTWARE: Wordperfect 6.0a for Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/840,077A
; FILING DATE: 20-FEB-1992
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/816,477
; FILING DATE: 03-JAN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Deborah A. Peacock
; REGISTRATION NUMBER: 31,649
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (505) 242-9677
; TELEFAX: (505) 243-2542
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-07-840-077A-7

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wk 2
||
Db 2 wk 3

RESULT 16
US-08-332-071B-16
; Sequence 16, Application US/08332071B
; Patent No. 5556836
; GENERAL INFORMATION:
; APPLICANT: ROEDERN, ERICH G.
; APPLICANT: KESSLER, HORST
; APPLICANT: KUTSCHER, BERNHARD
; APPLICANT: BERND, MICHAEL

APPLICANT: KLENNER, THOMAS
TITLE OF INVENTION: USE OF D-GLUCOPHRANDRONIC ACIDS AND
THEIR DERIVATIVES FOR INCORPORATION IN PHARMACOLOGICALLY
ACTIVE PEPTIDES AND THEIR SALTS
TITLE OF INVENTION: 17
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/332,071B
FILING DATE: 01-NOV-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: CHAPIN, MARLANA K.
REGISTRATION NUMBER: 35,843
REFERENCE/DOCKET NUMBER: 326/216933
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-332-071B-16

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 SY 6
11
Db 1 SY 2

RESULT 17
US-08-176-938-20
Sequence 20, Application US/08176938
Patent No. 5602099
GENERAL INFORMATION:
APPLICANT: Schiller, Peter W.
TITLE OF INVENTION: New Peptides
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: White & Case
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: U.S.A.
ZIP: 10036-2787
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/176,938
FILING DATE: 04-JAN-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:

NAME: Sterner Ph.D., Richard J.
REGISTRATION NUMBER: 35,372
REFERENCE/DOCKET NUMBER: 1103326-080
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-819-8783
TELEFAX: 212-354-8113
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 2
OTHER INFORMATION: /label= Tic
OTHER INFORMATION: /note= "1,2,3,4-tetrahydroisquinoline-3-carboxylic acid"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 3
OTHER INFORMATION: /label= Phg
OTHER INFORMATION: /note= "phenylglycine"
US-08-176-938-20

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 YXG 8
111
Db 1 YXG 3

RESULT 18
US-08-487-006-150
Sequence 150, Application US/08487006
Patent No. 5641861
GENERAL INFORMATION:
APPLICANT: Dooley, Colette T.
TITLE OF INVENTION: Houghten, Richard A.
TITLE OF INVENTION: No. 5641861e1 Mu Opioid Receptor Ligands:
NUMBER OF SEQUENCES: 222
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,006
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-TP 1706
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 150:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
TOPOLOGY: linear
FEATURE:

RESULT 21
US-08-487-006-153
Sequence 153, Application US/08487006
Patent No. 5641861
GENERAL INFORMATION:
APPLICANT: Dooley, Colette T.
APPLICANT: Houghten, Richard A.
TITLE OF INVENTION: No. 5641861el Mu Opioid Receptor Ligands:
TITLE OF INVENTION: Agonists and Antagonists
NUMBER OF SEQUENCES: 222
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,006
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-TP 1706
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 153:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
TOPOLOGY: linear
FEATURE:
NAME/KEY: Peptide
LOCATION: 2
OTHER INFORMATION: /note= "Xaa is (D)Nle."
FEATURE:
NAME/KEY: Peptide
LOCATION: 4
OTHER INFORMATION: /note= "amino acid is amidated at
OTHER INFORMATION: the C-terminal."
US-08-487-006-153
Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05; Indels 0; Gaps 0;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 6 YXG 8
DB 1 YXG 3
RESULT 22
US-08-170-360-17
Sequence 17, Application US/08170360
Patent No. 5656602
GENERAL INFORMATION:
APPLICANT: Tseng, Albert P. S.
APPLICANT: Inglis, Adam
TITLE OF INVENTION: PLA2 INHIBITORY COMPOUNDS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rothwell, Figg Ernst & Kurz
STREET: Suite 701-E, 555 Thirteenth St., N W
CITY: Washington
STATE: D. C.

COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/170,360
FILING DATE: 03-MAR-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PCT/AU92/00333
FILING DATE: 06-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PK 7058
FILING DATE: 04-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: Ernst, Barbara G.
REGISTRATION NUMBER: 30,377
REFERENCE/DOCKET NUMBER: 1871-104A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)783-6040
TELEFAX: (202)783-6031
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: both
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-170-360-17
Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05; Indels 0; Gaps 0;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 SY 6
DB 3 SY 4
RESULT 23
US-08-454-950-7
Sequence 7, Application US/08454950
Patent No. 5690905
GENERAL INFORMATION:
APPLICANT: Zamora, Paul O.
APPLICANT: Rhodes, Buck A.
TITLE OF INVENTION: Peptide-Metal Ion
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rhoad Incorporated
STREET: 4261 Balloon Park
CITY: Albuquerque
STATE: NM
COUNTRY: U.S.A.
ZIP: 87109-5802
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB
MEDIUM TYPE: Storage
COMPUTER: IBM PC/XT/AT, IBM PS/2 or
COMPUTER: compatibles
OPERATING SYSTEM: PC-DOS or MS-DOS
SOFTWARE: WordPerfect 6.0a for Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/454,950
FILING DATE: 31-MAY-1995

CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/840,077
FILING DATE: 20-FEB-1992
APPLICATION NUMBER: 07/816,477
FILING DATE: 03-JAN-1992
ATTORNEY/AGENT INFORMATION:
NAME: Deborah A. Peacock
REGISTRATION NUMBER: 31,649
TELECOMMUNICATION INFORMATION:
TELEPHONE: (505) 242-9677
TELEFAX: (505) 243-2542
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-454-950-7

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wk 2
11
DB 2 wk 3

RESULT 24
US-08-434-761-3
Sequence 3, Application US/08434761
Patent No. 5698673
GENERAL INFORMATION:
APPLICANT: Blondelle, Sylvie E.
APPLICANT: Plinilla, Clemencia
APPLICANT: Elchler, Jutta
APPLICANT: Houghten, Richard A.
TITLE OF INVENTION: PEPTIDES HAVING ANTI-MICROBIAL,
TITLE OF INVENTION: HEMOLYTIC, ANTI-MELITTIN OR ANTI-ENZYMATIC ACTIVITY
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESSES:
ADDRESSEE: CAMPBELL AND FLORES
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States of America
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/434,761
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/079,445
FILING DATE: 18-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-HP 9648
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:

LENGTH: 4 amino acids
TYPE: amino acid
TOPOLOGY: linear
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /label= Xaa
OTHER INFORMATION: /note= "Xaa-Kfmoc"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /label= Xaa
OTHER INFORMATION: /note= "Xaa-any amino acid"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /note= "C-terminal amino acid is
OTHER INFORMATION: amidated"
US-08-434-761-3

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wk 2
11
DB 2 wk 3

RESULT 25
US-08-338-890B-1
Sequence 1, Application US/08338890B
Patent No. 5700905
GENERAL INFORMATION:
APPLICANT: Hirschmann, Ralph
APPLICANT: Spanevello, Rolando
APPLICANT: Nutt, Ruth
TITLE OF INVENTION: Synthetic Somatostatin Mimics
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz and No. 5700905rls
STREET: One Liberty Place, 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS-DOS/WINDOWS
SOFTWARE: WORDPERECT SAVED AS ASCII (DOS) TEXT(*.*)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/338,890B
FILING DATE: NOV. 14, 1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Malinoski, Lynn
REGISTRATION NUMBER: 38,788
REFERENCE/DOCKET NUMBER: UPN-2040
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215)568-3100
TELEFAX: (215)568-3439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-338-890B-1
Query Match 25.0%; Score 2; DB 1; Length 4;

Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 wk 2
11
Db 2 WK 3

Search completed: January 14, 2002, 07:57:30
Job time: 375 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:58:36 ; Search time 45.4 Seconds
(without alignments)
13.423 Million cell updates/sec

Title: 09-185908-1b
Perfect score: 8
Sequence: 1 wxxxxxyxg 8

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

Database :

PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3	37.5	113	2 A41073	thyroglobulin - ra
2	2	25.0	5	2 E60274	major protein anti
3	2	25.0	5	2 S68326	blood cell protein
4	2	25.0	7	2 PC2132	FMRFamide-related
5	2	25.0	7	2 S33244	neuromodulatory pe
6	2	25.0	7	2 S33246	nitrate reductase
7	2	25.0	8	2 S68802	Ig heavy chain CRD
8	2	25.0	8	2 PT0311	quinoline 2-oxidor
9	2	25.0	9	2 S66607	gonadoliberin - se
10	2	25.0	10	1 RHLMGS	thrombospondin 2 -
11	2	25.0	10	2 C45474	polygalacturonase
12	2	25.0	10	2 S62880	gonadotropin-relea
13	2	25.0	10	2 A49187	Ig heavy chain CRD
14	2	25.0	10	2 PT0309	small nuclear ribo
15	2	25.0	10	2 I48778	protein QA300023 -
16	2	25.0	11	2 PT0081	ribosomal protein
17	2	25.0	11	2 S78026	urotensin II - lon
18	2	25.0	12	1 UOGM2	urotensin II - tel
19	2	25.0	12	2 S42765	urotensin II-A pep
20	2	25.0	12	2 JS0423	proton-translocati
21	2	25.0	12	2 S69123	photosystem II 3.7
22	2	25.0	12	2 S01122	acidic ribosomal p
23	2	25.0	12	2 PA0019	plastocyanin 2 - A
24	2	25.0	12	2 PA0037	ribosomal protein
25	2	25.0	12	2 PN0160	urotensin II-B pep
26	2	25.0	12	2 JS0424	T-cell receptor be
27	2	25.0	12	2 PH0930	amino transferase c
28	2	25.0	12	4 PC2122	urotensin II - lau
29	2	25.0	13	2 PQ0445	

30	2	25.0	13	2 S01119	photosystem II pro
31	2	25.0	13	2 PN0176	acidic ribosomal p
32	2	25.0	13	2 PQ0700	unidentified 6.3/4
33	2	25.0	13	2 A61514	glutathione transf
34	2	25.0	13	2 A60379	factor X activator
35	2	25.0	13	2 A33660	osteoclast functio
36	2	25.0	13	2 PT0290	Ig heavy chain CRD
37	2	25.0	13	2 S47368	T-cell antigen rec
38	2	25.0	13	2 S47372	T-cell antigen rec
39	2	25.0	13	2 S47384	T-cell antigen rec
40	2	25.0	13	2 B56864	dipeptidyl-peptida
41	2	25.0	13	2 PH0786	T-cell receptor al
42	2	25.0	13	2 S66558	serine proteinase
43	2	25.0	14	1 QMVRXX	mastoparan X - hor
44	2	25.0	14	1 QMVRPP	polistes mastopara
45	2	25.0	14	2 C60414	somatostatin - sil
46	2	25.0	14	2 B60842	somatostatin I - c
47	2	25.0	14	2 A60622	somatostatin - spo
48	2	25.0	14	2 A60840	somatostatin I - E
49	2	25.0	14	2 S00172	somatostatin I - s
50	2	25.0	14	2 JH0328	probursin tetradec

ALIGNMENTS

RESULT 1
A41073
thyroglobulin - rabbit (fragments)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C>Date: 12-Jun-1992 #sequence_revision 12-Jun-1992 #text_change 31-Oct-1997
C:Accession: A41073
R:Dunn, A.D.; Crutchfield, H.E.; Dunn, J.T.
J. Biol. Chem. 266, 20198-20204, 1991
A>Title: Thyroglobulin processing by thyroidal proteases. Major sites of cleavage by
A:Reference number: A41073; MUID:92041846
A:Accession: A41073
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-113 <DUN>
C:Superfamily: thyroglobulin; cholinesterase homology; thyroglobulin type I repeat ho

Query Match 37.5%; Score 3; DB 2; Length 113;
Best Local Similarity 100.0%; Pred. No. 6.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 ayxg 8
DB 64 AYXG 67

RESULT 2
E60274
major protein antigen MPT63 - Mycobacterium tuberculosis (fragment)
C:Species: Mycobacterium tuberculosis
C>Date: 11-Dec-1992 #sequence_revision 11-Dec-1992 #text_change 30-Sep-1993
C:Accession: E60274
R:Nagai, S.; Wiker, H.G.; Harboe, M.; Kinomoto, M.
Infect. Immun. 59, 372-382, 1991
A>Title: Isolation and partial characterization of major protein antigens in the cult
A:Reference number: A60274; MUID:91099989
A:Accession: E60274
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-5 <NAG>

Query Match 25.0%; Score 2; DB 2; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 ay 6

Db 1 AY 2

RESULT 3

S68326

Blood cell protein B - Ascidia ceratodes (fragment)

N:Alternate names: Abcp-B

C:Species: Ascidia ceratodes

C:Date: 19-Mar-1997 #sequence_revision 29-Aug-1997 #text_change 11-Jun-1999

C:Accession: S68326

R:Taylor, S.W.; Ross, M.M.; Waite, J.H.

Arch. Biochem. Biophys. 324, 228-240, 1995

A:Title: Novel 3,4-di- and 3,4,5-trihydroxyphenylalanine-containing polypeptides from th

A:Reference number: S68325; MUID:96132650

A:Accession: S68326

A:Molecule type: protein

A:Residues: 1-5 <TAY>

F:2/Modified site: 3',4',5'-trihydroxyphenylalanine (Tyr) #status experimental

F:4/Modified site: 3',4'-dihydroxyphenylalanine (Tyr) #status experimental

Query Match 25.0%; Score 2; DB 2; Length 5;

Best Local Similarity 100.0%; Pred. No. 2.2e+05;

Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6

Db 3 AY 4

RESULT 4

PC2132

PMRamide-related heptapeptide - Panagrellus redivivus

C:Species: Panagrellus redivivus

C:Date: 03-May-1994 #sequence_revision 15-Oct-1994 #text_change 11-Jul-1997

C:Accession: PC2132

R:Maule, A.G.; Shaw, C.; Bowman, J.W.; Halton, D.W.; Thompson, D.P.; Geary, T.G.; Thim,

Biochem. Biophys. Res. Commun. 200, 973-980, 1994

A:Title: KATYMRamide: a novel PMRamide-related heptapeptide from the free-living nemat

A:Reference number: PC2132; MUID:94235053

A:Accession: PC2132

A:Molecule type: protein

A:Residues: 1-7 <MAU>

C:Keywords: amidated carboxyl end

F:7/Modified site: amidated carboxyl end (Phe) #status predicted

Query Match 25.0%; Score 2; DB 2; Length 7;

Best Local Similarity 100.0%; Pred. No. 2.2e+05;

Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6

Db 3 AY 4

RESULT 5

S33244

neuromodulatory peptide Wamide-1 - giant African snail

C:Species: Achatina fulica (giant African snail)

C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Jul-1997

C:Accession: S33244

R:Minakata, H.; Ikeda, T.; Muneoka, Y.; Kobayashi, M.; Nomoto, K.

FEBS Lett. 323, 104-108, 1993

A:Title: Wamide-1, -2 and -3: novel neuromodulatory peptides isolated from ganglia of t

A:Reference number: S33244; MUID:93265912

A:Accession: S33244

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-7 <MIN>

Query Match 25.0%; Score 2; DB 2; Length 7;

Best Local Similarity 100.0%; Pred. No. 2.2e+05;

Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2

Db 1 WK 2

RESULT 6

S33246

neuromodulatory peptide Wamide-3 - giant African snail

C:Species: Achatina fulica (giant African snail)

C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Jul-1997

C:Accession: S33246

R:Minakata, H.; Ikeda, T.; Muneoka, Y.; Kobayashi, M.; Nomoto, K.

FEBS Lett. 323, 104-108, 1993

A:Title: Wamide-1, -2 and -3: novel neuromodulatory peptides isolated from ganglia o

A:Reference number: S33244; MUID:93265912

A:Accession: S33246

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-7 <MIN>

Query Match 25.0%; Score 2; DB 2; Length 7;

Best Local Similarity 100.0%; Pred. No. 2.2e+05;

Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2

Db 1 WK 2

RESULT 7

S68802

nitrate reductase (NADH) inhibitor - spinach (fragment)

C:Species: Spinacia oleracea (spinach)

C:Date: 04-Dec-1997 #sequence_revision 12-Dec-1997 #text_change 30-Jan-1998

C:Accession: S68802

R:Bachmann, M.; Huber, J.L.; Liao, P.C.; Gage, D.A.; Huber, S.C.

FEBS Lett. 387, 127-131, 1996

A:Title: The inhibitor protein of phosphorylated nitrate reductase from spinach (Spin

A:Reference number: S68802; MUID:96244508

A:Accession: S68802

A:Molecule type: protein

A:Residues: 1-8 <BAC>

A:Experimental source: leaves; strain cv. Bloomsdale

Query Match 25.0%; Score 2; DB 2; Length 8;

Best Local Similarity 100.0%; Pred. No. 2.2e+05;

Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6

Db 6 AY 7

RESULT 8

PT0311

Ig heavy chain CRD3 region (clone 6-100) - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Aug-1996

C:Accession: PT0311

R:Yamada, M.; Wasserman, R.; Reichard, B.A.; Shane, S.; Caton, A.J.; Rovera, G.

J. Exp. Med. 173, 395-407, 1991

A:Title: Preferential utilization of specific immunoglobulin heavy chain diversity an

A:Reference number: PT0222; MUID:91108337

A:Accession: PT0311

A:Molecule type: DNA

A:Residues: 1-8 <YAM>

A:Experimental source: B lymphocyte

C:Keywords: heterotetramer; immunoglobulin

Query Match 25.0%; Score 2; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6
||
DB 3 AY 4

RESULT 9
S6607

quinoline 2-oxidoreductase beta chain - Comamonas testosteroni (fragment)
C:Species: Comamonas testosteroni
C:Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 17-Mar-1999
C:Accession: S66607
R:Schach, S.; Tshisuaka, B.; Fetzner, S.; Lingens, F.
Eur. J. Biochem. 232, 536-544, 1995
A:Title: Quinoline 2-oxidoreductase and 2-oxo-1,2-dihydroquinoline 5,6-dioxygenase from
A:Reference number: S66606; MUID:96035889
A:Accession: S66607
A:Molecule type: protein
A:Residues: 1-9 <SCH>
A:Experimental source: strain 63

Query Match 25.0%; Score 2; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6
||
DB 7 AY 8

RESULT 10
RHLMGS

gonadoliberin - sea lamprey
N:Alternate names: gonadotropin releasing hormone (GnRH)
C:Species: Petromyzon marinus (sea lamprey)
C:Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 18-Mar-1997
C:Accession: A01412
R:Sherwood, N.M.; Sower, S.A.; Marshak, D.R.; Fraser, B.A.; Brownstein, M.J.
J. Biol. Chem. 261, 4812-4819, 1986
A:Title: Primary structure of gonadotropin-releasing hormone from lamprey brain.
A:Reference number: A01412; MUID:86168192
A:Accession: A01412
A:Molecule type: protein
A:Residues: 1-10 <SHE>

C:Comment: This hormone was isolated from the brain.
C:Superfamily: gonadoliberin
C:Keywords: amidated carboxyl end; hormone; pyroglutamic acid
F:10/Modified site: amidated carboxyl end (Gly) #status experimental
F:10/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 25.0%; Score 2; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
||
DB 7 WK 8

RESULT 11
C45474

thrombospondin 2 - bovine (fragment)
N:Alternate names: corticotropin-induced secreted protein (CISP); thrombospondin homolog
C:Species: Bos primigenius taurus (cattle)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 14-Aug-1998

C:Accession: C45474
R:Pellerin, S.; Lafeuillade, B.; Scherrer, N.; Gagnon, J.; Shi, D.L.; Chambaz, E.M.;
J. Biol. Chem. 268, 4304-4310, 1993
A:Title: Corticotropin-induced secreted protein, an ACTH-induced protein secreted by
A:Reference number: A45474; MUID:93179438
A:Accession: C45474
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-10 <PEL>

A:Experimental source: adrenocortical cells
A>Note: sequence extracted from NCBI backbone (NCBIP:125844)
C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology;

Query Match 25.0%; Score 2; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6
||
DB 1 AY 2

RESULT 12
S62880

polygalacturonase (EC 3.2.1.15) IV - Aspergillus sp. (fragment)
C:Species: Aspergillus sp.
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 17-Mar-1999
C:Accession: S62880
R:Straliova, E.; Dzurova, M.; Markovic, O.; Joernvall, H.
FEBS Lett. 382, 164-166, 1996
A:Title: An essential tyrosine residue of Aspergillus polygalacturonase.
A:Reference number: S62880; MUID:96196586
A:Accession: S62880
A:Molecule type: protein
A:Residues: 1-10 <STR>
C:Keywords: glycosidase; hydrolase
E:4/Active site: Tyr #status predicted

Query Match 25.0%; Score 2; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6
||
DB 3 AY 4

RESULT 13
A49187

gonadotropin-releasing hormone III - sea lamprey
C:Species: Petromyzon marinus (sea lamprey)
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 03-Mar-1995
C:Accession: A49187
R:Sower, S.A.; Chiang, Y.C.; Lovas, S.; Conlon, J.M.
Endocrinology 132, 1125-1131, 1993
A:Title: Primary structure and biological activity of a third gonadotropin-releasing
A:Reference number: A49187; MUID:93178316
A:Accession: A49187
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-10 <SOW>

A:Experimental source: brain
A>Note: sequence extracted from NCBI backbone (NCBIP:126381)

Query Match 25.0%; Score 2; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
||

Db 7 WK 8

RESULT 14

PT0309

Ig heavy chain CRD3 region (clone 6-94) - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Aug-1996

C:Accession: PT0309

R:Yamada, M.; Wasserman, R.; Reichard, B.A.; Shane, S.; Caton, A.J.; Rovera, G.

J. Exp. Med. 173, 395-407, 1991

A:Title: Preferential utilization of specific immunoglobulin heavy chain diversity and

A:Reference number: PT0222; MUID:91108337

A:Accession: PT0309

A:Molecule type: DNA

A:Residues: 1-10 <YAM>

A:Experimental source: B lymphocyte

C:Keywords: heterotetramer; immunoglobulin

Query Match 25.0%; Score 2; DB 2; Length 10;

Best Local Similarity 100.0%; Pred. No. 1.9e+03;

Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ay 6
||
Db 9 AY 10

RESULT 15

148778

small nuclear ribonucleoprotein E - mouse (fragment)

C:Species: Mus musculus (house mouse)

C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999

C:Accession: 148778

R:Fautsch, M.P.; Thompson, M.A.; Holicky, E.L.; Schultz, F.J.; Hallett, J.B.; Wieben, E.

Genomics 14, 883-890, 1992

A:Title: Conservation of coding and transcriptional control sequences within the snRNP E

A:Reference number: A4368; MUID:93122798

A:Accession: 148778

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-10 <RES>

A:Cross-references: EMBL.X65703; NID:g312006; PIDN:CAA46625.1; PID:g312007

Query Match 25.0%; Score 2; DB 2; Length 10;

Best Local Similarity 100.0%; Pred. No. 1.9e+03;

Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ay 6
||
Db 2 AY 3

RESULT 16

PT0081

protein QA300023 - Arabidopsis thaliana (fragment)

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 07-Feb-1996 #sequence_revision 19-Apr-1996 #text_change 24-Nov-1999

C:Accession: PT0081

R:Tsuigita, A.; Kamo, M.; Kawakami, M.; Ohki, Y.

submitted to JIPID, December 1995

A:Description: Two dimensional electrophoresis of plant proteins and standardization of

A:Reference number: PM0173

A:Accession: PT0081

A:Molecule type: protein

A:Residues: 1-11 <TSU>

A:Experimental source: Leaf

C:Keywords: acetylated amino end

F;1/Modified site: acetylated amino end (Ser) #status experimental

Query Match 25.0%; Score 2; DB 2; Length 11;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;QY 5 ay 6
||
Db 5 AY 6

RESULT 17

S78026

ribosomal protein YML29, mitochondrial - yeast (Saccharomyces cerevisiae) (fragment)

C:Species: Saccharomyces cerevisiae

C:Date: 09-Oct-1997 #sequence_revision 24-Oct-1997 #text_change 14-Nov-1997

C:Accession: S78026

R:Kitakawa, M.; Graack, H.R.; Grochmann, L.; Goldschmidt-Reisin, S.; Herfurth, E.; Wit

Eur. J. Biochem. 245, 449-456, 1997

A:Title: Identification and characterization of the genes for mitochondrial ribosomal

A:Reference number: S78018; MUID:97296414

A:Accession: S78026

A:Molecule type: protein

A:Residues: 1-11 <KIT>

C:Genetics:

A:Genome: nuclear

C:Keywords: mitochondrion; protein biosynthesis; ribosome

Query Match 25.0%; Score 2; DB 2; Length 11;

Best Local Similarity 100.0%; Pred. No. 2e+03;

Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ay 6
||
Db 4 AY 5

RESULT 18

U06M2

urotensin II - long-jawed mudsucker

C:Species: Gilllichthys mirabilis (long-jawed mudsucker)

C:Date: 23-Oct-1981 #sequence_revision 23-Oct-1981 #text_change 15-Oct-1996

C:Accession: A01409

R:Pearson, D.; Shively, J.E.; Clark, B.R.; Geschwind, I.I.; Barkley, M.; Nishio, R.

Proc. Natl. Acad. Sci. U.S.A. 77, 5021-5024, 1980

A:Title: Urotensin II: a somatostatin-like peptide in the caudal neurosecretory system

A:Reference number: A01409; MUID:81054904

A:Accession: A01409

A:Molecule type: protein

A:Residues: 1-12 <PEA>

A:Note: the proposed sequence was confirmed by synthesis of a peptide with the same s

C:Superfamily: urotensin II

C:Keywords: neuropeptide; osmoregulation

F;6-11/Disulfide bonds: #status experimental

Query Match 25.0%; Score 2; DB 1; Length 12;

Best Local Similarity 100.0%; Pred. No. 2.2e+03;

Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wk 2
||
Db 8 WK 9

RESULT 19

S42765

urotensin II - teleostean fish

C:Species: teleostean fish

C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 05-Dec-1997

C:Accession: S42765

R:Bhaskaran, R.; Arunkumar, A.I.; Yu, C.

Biochim. Biophys. Acta 1199, 115-122, 1994

A:Title: NMR and dynamical simulated annealing studies on the solution conformation of u
A:Reference number: S42765; MUID:94169160
A:Accession: S42765
A:Molecule type: protein
A:Residues: 1-12 <BHA>
C:Superfamily: urotensin II
C:Keywords: neuropeptide; osmoregulation
F:6-11/Product: urotensin II #status experimental <MAT>

Query Match 25.0%; Score 2; DB 2; Length 12;
Best Local Similarity 100.0%; Pred. No. 2.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
||
DB 8 wk 9

RESULT 20

JS0423
urotensin II-A peptide - white sucker
C:Species: Catostomus commersoni (white sucker)
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-May-1997
C:Accession: JS0423
R:McMaster, D.; Lederis, K.
Peptides 4, 367-373, 1983
A:Title: Isolation and amino acid sequence of two urotensin II peptides from Catostomus
A:Reference number: JS0423; MUID:84041959
A:Accession: JS0423
A:Molecule type: protein
A:Residues: 1-12 <MC>
C:Comment: This peptide has smooth muscle-stimulating activity.
C:Superfamily: urotensin II
F:6-11/Disulfide bonds: #status experimental

Query Match 25.0%; Score 2; DB 2; Length 12;
Best Local Similarity 100.0%; Pred. No. 2.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
||
DB 8 wk 9

RESULT 21

S69123
proton-translocating transhydrogenase - Rhodospirillum rubrum (fragment)
C:Species: Rhodospirillum rubrum
C:Date: 10-Mar-1998 #sequence_revision 24-Apr-1998 #text_change 24-Apr-1998
C:Accession: S69123
R:Diggle, C.; Hutton, M.; Jones, G.R.; Thomas, C.M.; Jackson, J.B.
Eur. J. Biochem. 228, 719-726, 1995
A:Title: Properties of the soluble polypeptide of the proton-translocating transhydrogen
A:Reference number: S69123; MUID:95255277
A:Accession: S69123
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-12 <DIG>

Query Match 25.0%; Score 2; DB 2; Length 12;
Best Local Similarity 100.0%; Pred. No. 2.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
||
DB 6 wk 7

RESULT 22

S01122

photosystem II 3.7k protein - spinach (fragment)
C:Species: Spinacia oleracea (spinach)
C:Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 18-Jun-1993
C:Accession: S01122
R:Schoeder, W.P.; Henrysson, T.; Akerlund, H.E.
FEBS Lett. 235, 289-292, 1988
A:Title: Characterization of low molecular mass proteins of photosystem II by N-termi
A:Reference number: S01120
A:Accession: S01122
A:Molecule type: protein
A:Residues: 1-12 <SCH>
C:Keywords: chloroplast; photosynthesis; photosystem II; thylakoid

Query Match 25.0%; Score 2; DB 2; Length 12;
Best Local Similarity 100.0%; Pred. No. 2.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6
||
DB 5 ay 6

RESULT 23

PA0019
acidic ribosomal P2-like protein - Arabidopsis thaliana (fragment)
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Jun-1992 #sequence_revision 06-Jan-1995 #text_change 11-Apr-1995
C:Accession: PA0019
R:Kamo, M.; Kawakami, T.; Miyatake, N.; Tsugita, A.
Submitted to JIPID, July 1994
A:Description: Separation and characterization of Arabidopsis proteins by two-dimensi
A:Reference number: PA0001
A:Accession: PA0019
A:Molecule type: protein
A:Residues: 1-12 <KAM>
A:Experimental source: callus

Query Match 25.0%; Score 2; DB 2; Length 12;
Best Local Similarity 100.0%; Pred. No. 2.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6
||
DB 6 ay 7

RESULT 24

PA0037
plastocyanin 2 - Arabidopsis thaliana (fragment)
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Jun-1992 #sequence_revision 06-Jan-1995 #text_change 23-Mar-1995
C:Accession: PA0037
R:Kamo, M.; Kawakami, T.; Miyatake, N.; Tsugita, A.
Submitted to JIPID, July 1994
A:Description: Separation and characterization of Arabidopsis proteins by two-dimensi
A:Reference number: PA0001
A:Accession: PA0037
A:Molecule type: protein
A:Residues: 1-12 <KAM>
A:Experimental source: stem

Query Match 25.0%; Score 2; DB 2; Length 12;
Best Local Similarity 100.0%; Pred. No. 2.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6
||
DB 9 ay 10

RESULT 25

PN0160

Ribosomal protein S16 - fungus (Fusarium sporotrichioides) (fragment)

C;Species: Fusarium sporotrichioides

C;Date: 05-Aug-1994 #sequence_revision 05-Aug-1994 #text_change 11-Nov-1994

C;Accession: PN0160

R;Fukaya, N.; Chow, L.P.; Sugiyura, Y.; Tsugita, A.; Ueno, Y.; Tabuchi, K.

submitted to JRPD, May 1994

A;Description: Two dimensional polyacrylamide gel electrophoresis of Fusarium sporotrich

A;Reference number: PN0160

A;Accession: PN0160

A;Molecule type: protein

A;Residues: 1-12 <PUK>

C;keywords: protein biosynthesis; ribosome

Query Match 25.0%; Score 2; DB 2; Length 12;

Best Local Similarity 100.0%; Pred.No. 2.2e+03;

Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ay 6

11

1 AY 2

Search completed: January 14, 2002, 07:58:36
Job time: 386 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 08:08:21 ; Search time 30.66 Seconds
(without alignments)
9.567 Million cell updates/sec

Title: 09-185908-1b

Perfect score: 8

Sequence: 1 wxxxxxxg 8

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size: 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

Database: SWISSprot_39.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	37.5	80	VA0H_BOVIN	P81103 bos taurus
2	25.0	6	OVN_LEPDE	P42985 leptoletars
3	25.0	7	ALL2_CAMA	P81805 carcius ma
4	25.0	7	PAR3_HAECO	P81298 haemochus
5	25.0	7	FAR3_PANRE	P41874 panagrellus
6	25.0	7	MNP1_LEPDE	P42984 leptoletars
7	25.0	7	WMA2_ACHFU	P35920 achalina fu
8	25.0	7	WMA3_ACHFU	P35921 achalina fu
9	25.0	8	UF06_MOUSE	P38644 mus musculu
10	25.0	10	GON1_PETMA	P04378 petromyzon
11	25.0	10	GON3_PETMA	P30948 petromyzon
12	25.0	10	Q20B_COMTE	P80465 comamonas t
13	25.0	12	PSP3_PHYPA	P80662 physcomitre
14	25.0	12	UR2A_CATCO	P04558 catostomus
15	25.0	12	UR2B_CATCO	P04559 catostomus
16	25.0	12	UR2B_CYPCA	P04561 cyprinus ca
17	25.0	12	UR2_GILMI	P01147 gillichthys
18	25.0	12	UR2_POLSP	P81022 polyodon sp
19	25.0	12	UR2_SCYCA	P35490 scyllorhinu
20	25.0	13	PSBP_PINPS	P81668 pinus pinus
21	25.0	13	UVRD_SATY	Q05311 salmoneilla
22	25.0	14	ECDC_LYMDI	P80940 lymphocitri
23	25.0	14	MAST_PARDI	P42716 parapolybia
24	25.0	14	MAST_POLJA	P01517 polistes ja
25	25.0	14	MAST_VESXA	P01515 vespa xanth
26	25.0	14	SMS1_MYOSC	P20750 myosepalm
27	25.0	14	SMS_ALLMI	P31885 alligator m
28	25.0	15	MALT_BACTQ	P80072 bacillus th
29	25.0	15	SODM_ENTAE	P22799 enterobacte
30	25.0	15	UC14_MAIZE	P80620 zea mays (m
31	25.0	15	VAA3_RHOPA	Q02006 rhodospheud
32	25.0	16	ALRX_PSEPU	P17916 pseudomonas
33	25.0	16	ARCD_PSEPU	P41147 pseudomonas

34	2	25.0	16	1	UPAB_HUMAN	P31935 homo saplen
35	2	25.0	17	1	ATP1_PAVLU	P28529 pavlova lut
36	2	25.0	17	1	GAST_MACMO	P33714 macaca mula
37	2	25.0	17	1	TPIS_PINPS	P81666 pinus pinus
38	2	25.0	17	1	VPFK_SALTY	Q92145 salmoneilla
39	2	25.0	18	1	AGI_EUPMA	P33889 euphorbia m
40	2	25.0	18	1	ALL2_CYPPO	P82153 cydia pomon
41	2	25.0	18	1	UC03_MAIZE	P80609 zea mays (m
42	2	25.0	20	1	CAT4_FASHE	P80528 fasciola he
43	2	25.0	20	1	LPP2_HUMAN	P56642 homo saplen
44	2	25.0	20	1	MCRG_METTE	P22950 methanosarc
45	2	25.0	20	1	RLC1_HALMA	P12740 halobacula
46	2	25.0	20	1	SYR_RAT	P40329 rattus norv
47	2	25.0	21	1	MDH_BURCE	P80537 burkholderi
48	2	25.0	21	1	MDH_PSEIN	P80538 pseudomonas
49	2	25.0	21	1	TERT_APIME	P56587 apis mellif
50	2	25.0	22	1	RL41_METVA	P54025 methanococc

ALIGNMENTS

```
RESULT 1
1
VA0H_BOVIN
ID VA0H_BOVIN STANDARD: PRT: 80 AA.
AC P81103; 018981;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE VACUOLAR ATP SYNTHASE SUBUNIT H (EC 3.6.1.34) (V-ATPASE H SUBUNIT)
DE (VACUOLAR PROTON PUMP H SUBUNIT) (V-ATPASE M9.2 SUBUNIT) (V-ATPASE
DE 9.2 KDA MEMBRANE ACCESSORY PROTEIN).
GN ATP6H.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
ON [1]
RP SEQUENCE OF 1-19, AND SEQUENCE OF 19-80 FROM N.A.
RC TISSUE=Adrenal medulla;
RX MEDLINE=98225166; PubMed=9556572;
RA Ludwig J., Kersch S., Brandt U., Pfeiffer K., Getlavi F., Apps D.K.,
RT "Identification and characterization of a novel 9.2-kDa membrane
RT sector-associated protein of vacuolar proton-ATPase from chromaffin
RT granules."
RT J. Biol. Chem. 273:10939-10947(1998).
CC -!- FUNCTION: VACUOLAR ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY
CC -!- OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS.
CC -!- SUBUNIT: COMPOSED OF AT LEAST 10 SUBUNITS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN LUNG, HEART, SPLEEN, KIDNEY AND
CC ADRENAL GLAND. NOT FOUND IN BRAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: Y15285; CA75570.1; -.
CC Hydrolyase; Hydrogen ion transport; Transmembrane.
CC TRANSMEM 7 27
CC TRANSMEM 35 55
CC SEQUENCE 80 AA; 9129 MW; 87B2CEB7D47E5427 CRC64;
Query Match 37.5%; Score 3; DB 1; Length 80;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 5 ayxg 8
    ||||
Db 1 AYXG 4

RESULT 2
OVM_LEPDE
ID OVM_LEPDE STANDARD; PRT: 6 AA.
AC P42985;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE OVIDUCTAL MOTILITY STIMULATING PEPTIDE (LED-OVM).
OC Leptinotarsa decemlineata (Colorado potato beetle);
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Phytophaga; Chrysomeloidea; Chrysomelidae;
OC Chrysomelinae; Leptinotarsa.
OX NCBI_TaxID=7539;

RESULT 3
ALL2_CARMA
ID ALL2_CARMA STANDARD; PRT: 7 AA.
AC P81805;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CARINUSTATIN 2.
OS Carcinus maenas (Common shore crab) (Green crab).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC Eubrachyura; Fortuinidea; Portunidae; Carcinus.
OX NCBI_TaxID=6759;

RESULT 4
TISSEU-Head;
MEDLINE=91271080; PubMed=2052497;
RA Spittaels K., Schoofs L., Grauwels L., Smet H., van Damme J.,
RA Proost P., Torrekens S., de Loof A.;
RT "Isolation, identification and synthesis of novel oviductal motility
stimulating head peptide in the Colorado potato beetle, Leptinotarsa
decemlineata."
RL Peptides 12:31-36(1991).
CC -I- FUNCTION: MYOTROPIC PEPTIDE. STIMULATES THE CONTRACTIONS OF THE
OVIDUCT.
CC Neuropeptide; Amidation.
FT MOD.RES 6 6 AMIDATION.
SQ SEQUENCE 6 AA; 720 MW; 6B07632B5DD03000 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 6;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6
    ||
Db 2 AY 3
```

```
Query Match 25.0%; Score 2; DB 1; Length 7;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6
    ||
Db 2 AY 3

RESULT 4
FAR3_HAECCO
ID FAR3_HAECCO STANDARD; PRT: 7 AA.
AC P81298;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE FMRFAMIDE-LIKE NEUROPEPTIDE PF3 (KSAYMRF-AMIDE).
OS Haemonchus contortus
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
OC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.
OX NCBI_TaxID=6289;

RESULT 5
FAR3_PANRE
ID FAR3_PANRE STANDARD; PRT: 7 AA.
AC P41874;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE FMRFAMIDE-LIKE NEUROPEPTIDE PF3 (KSAYMRF-AMIDE).
OS Panagrellus redivivus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
OC Panagrolaimoidea; Panagrolaimidae; Panagrellus.
OX NCBI_TaxID=6235;

RESULT 6
FAR3_PANRE
ID FAR3_PANRE STANDARD; PRT: 7 AA.
AC P41874;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE FMRFAMIDE-LIKE NEUROPEPTIDE PF3 (KSAYMRF-AMIDE).
OS Panagrellus redivivus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
OC Panagrolaimoidea; Panagrolaimidae; Panagrellus.
OX NCBI_TaxID=6235;

SEQUENCE, AND SYNTHESIS.
RX MEDLINE=94235053; PubMed=8179635;
RA Maule A.G., Shaw C., Bowman J.W., Halton D.W., Thompson D.P.,
RA Geary T.G., Thim L.;
RT "KSAYMRFamide: a novel FMRFamide-related heptapeptide from the free-
living nematode, Panagrellus redivivus, which is myoactive in the
RT parasitic nematode, Ascaris suum."
RL Biochem. Biophys. Res. Commun. 200:973-980(1994).
CC -I- FUNCTION: MYOACTIVE; INDUCES A RAPID CONCENTRATION-DEPENDENT
MUSCLE TENSION INCREASE.
CC -I- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
FAMILY.
DR PIR: PC2132; PC2132.
KW Neuropeptide; Amidation.
```

FT MOD_RES 7 7 AMIDATION.
SQ SEQUENCE 7 AA: 902 MW: 69d4068b5dc5b350 CRC64:

Query Match 25.0%; Score 2; DB 1; Length 7;
Best Local Similarity 100.0%; Pred. No. 1e+05; Mismatches 0; Indels 0; Gaps 0;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6
11
Db 3 AY 4

RESULT 6

MNP1_LEPDE
ID MNP1_LEPDE STANDARD: PRT: 7 AA.
AC P42984:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE MYOTROPIC NEUROPEPTIDE 1 (LED-MNP-1).
OC Leptinotarsa decemlineata (Colorado potato beetle).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Phycophaga; Chrysomelidae; Chrysomelidae;
OC Chrysomelinae; Leptinotarsa.
OX NCBI_TaxID=7539;
RN [1]
RP SEQUENCE, AND SYNTHESIS.
RC TISSUE-Head:
RX MEDLINE=93380343; PubMed=7651886;
RA Splitzels K., Vankeerberghen A., Schoofs L., Torrekens S.,
RA Grauwels L., van Leuven F., de Loof A.;
RT "Identification, characterization, and immunological localization of
RT a novel myotropic neuropeptide in the Colorado potato beetle,
RL Leptinotarsa decemlineata.";
RL Peptides 16:365-374(1995).
CC OVIDUCT.
CC Neuropeptide; Amidation.
FT MOD_RES 7 7 AMIDATION.
SQ SEQUENCE 7 AA: 705 MW: 6dd73768745b5db0 CRC64:

Query Match 25.0%; Score 2; DB 1; Length 7;
Best Local Similarity 100.0%; Pred. No. 1e+05; Mismatches 0; Indels 0; Gaps 0;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6
11
Db 1 AY 2

RESULT 7

WM22_ACHFV
ID WM22_ACHFV STANDARD: PRT: 7 AA.
AC P35920:
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE WM22-2.
OS Achatina fulica (Giant African snail).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;
OC Achatinacea; Achatinidae; Achatina.
OX NCBI_TaxID=6530;
RN [1]
RP SEQUENCE.
RC TISSUE-Ganglion;
RX MEDLINE=93265912; PubMed=8495720;
RA Minakata H., Ikeda T., Muneoka Y., Kobayashi M., Nomoto K.;
RT "Wamide-1, -2 and -3: novel neuromodulatory peptides isolated from
RT ganglia of the African giant snail, Achatina fulica.";
RL FEBS Lett. 323:104-108(1993).

KW Neuropeptide; Amidation.
FT MOD_RES 7 7 AMIDATION.
SQ SEQUENCE 7 AA: 964 MW: 7362d5b686d32310 CRC64:

Query Match 25.0%; Score 2; DB 1; Length 7;
Best Local Similarity 100.0%; Pred. No. 1e+05; Mismatches 0; Indels 0; Gaps 0;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
11
Db 1 WK 2

RESULT 8

WM23_ACHFV
ID WM23_ACHFV STANDARD: PRT: 7 AA.
AC P35921;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE WM23-3.
OC Achatina fulica (Giant African snail).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;
OC Achatinacea; Achatinidae; Achatina.
OX NCBI_TaxID=6530;
RN [1]
RP SEQUENCE.
RC TISSUE-Ganglion;
RX MEDLINE=93265912; PubMed=8495720;
RA Minakata H., Ikeda T., Muneoka Y., Kobayashi M., Nomoto K.;
RT "Wamide-1, -2 and -3: novel neuromodulatory peptides isolated from
RT ganglia of the African giant snail, Achatina fulica.";
RL FEBS Lett. 323:104-108(1993).
DR PIR: S33244; S33244.
KW Neuropeptide; Amidation.
FT MOD_RES 7 7 AMIDATION.
SQ SEQUENCE 7 AA: 965 MW: 7362d5b686d32310 CRC64:

Query Match 25.0%; Score 2; DB 1; Length 7;
Best Local Similarity 100.0%; Pred. No. 1e+05; Mismatches 0; Indels 0; Gaps 0;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
11
Db 1 WK 2

RESULT 9

UF06_MOUSE
ID UF06_MOUSE STANDARD: PRT: 8 AA.
AC P38644;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE UNKNOWN PROTEIN FROM 2D-PAGE OF FIBROBLASTS (P50) (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE.
RC TISSUE-Fibroblast;
RX MEDLINE=95009907; PubMed=7523108;
RA Merrick B.A., Patterson R.M., Wichter L.L., He C., Selkirk J.K.;
RT "Separation and sequencing of familial and novel murine proteins
RT using preparative two-dimensional gel electrophoresis.";
RL Electrophoresis 15:735-745(1994).
CC -1- MISCELLANEOUS: ON THE 2D-GEL, THE DETERMINED PI OF THIS UNKNOWN
CC PROTEIN IS: 5.2, ITS MW IS: 50 KDA.
FT NON_TER 8 8
SQ SEQUENCE 8 AA: 817 MW: A35D878676B05B1 CRC64:

```

Query Match          25.0%; Score 2; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6
    11
Db 7 At 8

RESULT 10
GONI_PETMA STANDARD; PRT; 10 AA.
ID GONI_PETMA
AC P04378:
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE GONADOLIBERIN I (GONADOTROPIN-RELEASING HORMONE I) (GNRH-I)
DE (LULIBERIN I).
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
OX NCBI_TaxID=7757;
RN [1]
RP SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=86168192; PubMed=3514603;
RA Sherwood N.M., Sower S.A., Marshak D.R., Fraser B.A., Brownstein M.J.;
RT "Primary structure of gonadotropin-releasing hormone from lamprey
   brain.";
RL J. Biol. Chem. 261:4812-4819(1986).
CC -1- FUNCTION: STIMULATES THE SECRETION OF BOTH LUTEINIZING AND
   FOLLICLE-STIMULATING HORMONES.
CC -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.
DR PIR: A01412; RHLMS.
DR InterPro: IPR002012; GNRH.
DR Pfam: PF00446; GNRH; 1.
DR PROSITE: PS00473; GNRH; 1.
KW Hormone; Amidation; Hypothalamus.
FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 10 10 AMIDATION.
SQ SEQUENCE 10 AA; 1244 MW; 1E4B36237B1735AB CRC64;

Query Match          25.0%; Score 2; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 8.4e+02;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
    11
Db 7 Wk 8

RESULT 11
GONI_PETMA STANDARD; PRT; 10 AA.
ID GONI_PETMA
AC P30948:
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GONADOLIBERIN III (GONADOTROPIN-RELEASING HORMONE III) (GNRH-III)
DE (LULIBERIN III).
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
OX NCBI_TaxID=7757;
RN [1]
RP SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=93178316; PubMed=8440174;
RA Sower S.A., Chiang Y.-C., Lovas S., Conlon J.M.;
RT "Primary structure and biological activity of a third gonadotropin-
```

```

RT releasing hormone from lamprey brain.";
RL Endocrinology 132:1125-1131(1993).
CC -1- FUNCTION: STIMULATES THE SECRETION OF BOTH LUTEINIZING AND
   FOLLICLE-STIMULATING HORMONES.
CC -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.
DR InterPro: IPR002012; GNRH.
DR Pfam: PF00446; GNRH; 1.
DR PROSITE: PS00473; GNRH; 1.
KW Hormone; Amidation; Hypothalamus.
FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 10 10 AMIDATION.
SQ SEQUENCE 10 AA; 1277 MW; 284B36237A1F5A3 CRC64;

Query Match          25.0%; Score 2; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 8.4e+02;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
    11
Db 7 Wk 8

RESULT 12
Q20B_COMTE STANDARD; PRT; 10 AA.
ID Q20B_COMTE
AC P80465:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE QUINOLINE 2-OXIDOREDUCTASE, BETA CHAIN (EC 1.3.99.17) (FRAGMENT).
OS Comamonas testosteroni (Pseudomonas testosteroni).
CC Bacteria; Proteobacteria; beta subdivision; Comamonadaceae; Comamonas.
OX NCBI_TaxID=285;
RN [1]
RP SEQUENCE.
RC STRAIN=63;
RX MEDLINE=96035889; PubMed=7556204;
RA Schach S., Tshisaka B., Fetzner S., Jiggins F.;
RT dioxygenase from Comamonas testosteroni 63. The first two enzymes in
   quinaline and 3-methylquinaline degradation.";
RL Eur. J. Biochem. 232:536-544(1995).
CC -1- FUNCTION: CONVERTS (3-METHYL-)-QUINOLINE TO (3-METHYL-)-2-OXO-
   1,2-DIHYDROQUINOLINE.
CC -1- CATALYTIC ACTIVITY: QUINOLINE + ACCEPTOR + H(2)O -> ISOQUINOLIN-
   1(2H)-ONE + REDUCED ACCEPTOR.
CC -1- COFACTOR: FAD; MOLYBDENUM AND IRON-SULFUR.
CC -1- PATHWAY: FIRST STEP IN THE DEGRADATION OF QUINOLINE AND
   (3-METHYL-)-QUINOLINE.
CC -1- SUBUNIT: HETEROHEXAMER OF TWO ALPHA CHAINS, TWO BETA CHAINS, AND
   TWO GAMMA CHAINS (PROBABLE).
KW Oxidoreductase; Flavoprotein; FAD; Molybdenum.
FT NON_TER 10 10
FT NON_TER 10 10
SQ SEQUENCE 10 AA; 1241 MW; C2E2C25DD9CDC769 CRC64;

Query Match          25.0%; Score 2; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 8.4e+02;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6
    11
Db 7 At 8

RESULT 13
PSP3_PHYPA STANDARD; PRT; 12 AA.
ID PSP3_PHYPA
AC P80662:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
```

DE OXYGEN-EVOLVING ENHANCER PROTEIN 2 (OE2) (24 KDA SUBUNIT OF OXYGEN
DE EVOLVING SYSTEM OF PHOTOSYSTEM II) (FRAGMENT).
OS Physcomitrella patens (Moss).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
OC Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
ON NCBI_TaxID=3218;
RN [1]
RP SEQUENCE.
RX TISSUE=Protonema; PubMed=9129336;
RC MEDLINE=97275459; PubMed=9129336;
RA Kaestn B., Buck F., Nuske J., Reski R.;
RT "Cytoklnh affects nuclear- and plastome-encoded energy-converting
Plastid enzymes";
PL Planta 201;261-272(1997).
RL -1- FUNCTION: MAY BE INVOLVED IN THE REGULATION OF PHOTOSYSTEM II.
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST THYLAKOID MEMBRANE; ASSOCIATED
WITH THE PHOTOSYSTEM II COMPLEX.
CC -1- INDUCTION: BY LIGHT.
CC -1- SIMILARITY: BELONGS TO THE PSBP FAMILY.
CC Photosynthesis; Photosystem II; Chloroplast; Thylakoid; Membrane;
KM Multigene family.
FT NON_TER 12
SQ SEQUENCE 12 AA: 1182 MW; 8D2BD54D7C44DC5 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 9.8e+02;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6
||
DB 1 AY 2

RESULT 14
UR2A_CATCO ID UR2A_CATCO STANDARD; PRT: 12 AA.
AC P04558;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE UROTENSIN IIA (U-IIA) (UIIA).
OS Catostomus commersoni (White sucker).
OC Catostomus commersoni (White sucker).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes;
OC Cypriniformes; Catostomidae; Catostomus.
ON NCBI_TaxID=7971;
RN [1]
RP SEQUENCE.
RX MEDLINE=84041959; PubMed=6138758;
RA McMaster D., Lederis K.;
RT "Isolation and amino acid sequence of two urotensin II peptides from
Catostomus commersoni urophyses.";
RL Peptides 4:367-373(1983).
CC -1- FUNCTION: UROTENSIN IS FOUND IN THE TELEOST CAUDAL NEUROSECRETORY
SYSTEM. IT HAS A SUGGESTED ROLE IN OSMOREGULATION AND AS A
CORTICOTROPIN-RELEASING FACTOR.
CC -1- SIMILARITY: BELONGS TO THE UROTENSIN 2 FAMILY.
CC PIR: JS0423; JS0423.
DR InterPro: IPR001483; Urotensin_II.
DR Pfam: PF02083; Urotensin_II; 1.
DR PROSITE: PS00984; UROTENSIN_II; 1.
KM Hormone.
FT DISULFID 6
SQ SEQUENCE 12 AA: 1336 MW; 969C76DBB879CEBA CRC64;

Query Match 25.0%; Score 2; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 9.8e+02;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 wk 2
||

DB 8 WK 9
RESULT 15
UR2B_CATCO ID UR2B_CATCO STANDARD; PRT: 12 AA.
AC P04559;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE UROTENSIN IIB (U-IIB) (UIIB).
OS Catostomus commersoni (White sucker).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes;
OC Cypriniformes; Catostomidae; Catostomus.
ON NCBI_TaxID=7971;
RN [1]
RP SEQUENCE.
RX MEDLINE=84041959; PubMed=6138758;
RA McMaster D., Lederis K.;
RT "Isolation and amino acid sequence of two urotensin II peptides from
Catostomus commersoni urophyses.";
RL Peptides 4:367-373(1983).
CC -1- FUNCTION: UROTENSIN IS FOUND IN THE TELEOST CAUDAL NEUROSECRETORY
SYSTEM. IT HAS A SUGGESTED ROLE IN OSMOREGULATION AND AS A
CORTICOTROPIN-RELEASING FACTOR.
CC -1- SIMILARITY: BELONGS TO THE UROTENSIN 2 FAMILY.
CC PIR: JS0424; JS0424.
DR InterPro: IPR001483; Urotensin_II.
DR Pfam: PF02083; Urotensin_II; 1.
DR PROSITE: PS00984; UROTENSIN_II; 1.
KM Hormone.
FT DISULFID 6
SQ SEQUENCE 12 AA: 1437 MW; 73961BDBB879CEBB CRC64;

Query Match 25.0%; Score 2; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 9.8e+02;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
||
DB 8 WK 9

RESULT 16
UR2B_CYPCA ID UR2B_CYPCA STANDARD; PRT: 12 AA.
AC P04561;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE UROTENSIN II-BETA.
OS Cyprinus carpio (Common carp).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes;
OC Cypriniformes; Cyprinidae; Cyprininae; Cyprinus.
ON NCBI_TaxID=7962;
RN [1]
RP SEQUENCE.
RX Munekata E., Ohtaki T., Ichikawa T., McMaster D., Lederis K.;
RT (In) Rich D.H., Gross E. (eds.);
Proceedings of the 7th american peptide symposium, pp. 69-72,
RL Pierce Chemical Co., Rockford IL. (1981).
CC -1- FUNCTION: UROTENSIN IS FOUND IN THE TELEOST CAUDAL NEUROSECRETORY
SYSTEM. IT HAS A SUGGESTED ROLE IN OSMOREGULATION AND AS A
CORTICOTROPIN-RELEASING FACTOR.
CC -1- SIMILARITY: BELONGS TO THE UROTENSIN 2 FAMILY.
CC InterPro: IPR001483; Urotensin_II.
DR Pfam: PF02083; Urotensin_II; 1.
DR PROSITE: PS00984; UROTENSIN_II; 1.
KM Hormone.
FT DISULFID 6
SQ SEQUENCE 12 AA: 1437 MW; 73961BDBB879CEBB CRC64;

FT VARIANT 2 2 G->S.
SQ SEQUENCE 12 AA: 1407 MW: 73960A9FB879CEBB CRC64;
Query Match 25.0%; Score 2; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 9.8e+02;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 wk 2
DB 8 WK 9
RESULT 17
UR2_GILMT STANDARD; PRT; 12 AA.
ID UR2_GILMT
AC P01147;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE UROTENSIN_II (U-II) (UII).
OS Gillichthys mirabilis (Long-jawed mudsucker).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Perciformes; Gobioidel;
OC Gobiidae; Gillichthys.
OX NCBI_TaxID=8222;
RN [1]
RP SEQUENCE.
RX MEDLINE=81054904; PubMed=6107911;
RA Pearson D., Shively J.E., Clark B.R., Geschwind I.I., Barkley M.,
Nishio K.R., Bern H.A.;
RT "Urotensin II: a somatostatin-like peptide in the caudal
neurosecretory system of fishes";
RL Proc. Natl. Acad. Sci. U.S.A. 77:5021-5024(1980).
DE -1- FUNCTION: UROTENSIN IS FOUND IN THE TELEOST CAUDAL NEUROSECRETORY
SYSTEM. IT HAS A SUGGESTED ROLE IN OSMOREGULATION AND AS A
CORTICOTROPIN-RELEASING FACTOR.
CC -1- SIMILARITY: BELONGS TO THE UROTENSIN 2 FAMILY.
DR PIR: A01409; UOGW2.
DR InterPro: IPR001483; Urotensin_II.
DR Pfam: PF02083; Urotensin_II; 1.
DR PROSITE: PS00984; UROTENSIN_II; 1.
KW Hormone.
FT DISULFID 6 11
SQ SEQUENCE 12 AA: 1364 MW: 968BF8982679CEBA CRC64;
Query Match 25.0%; Score 2; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 9.8e+02;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 wk 2
DB 8 WK 9
RESULT 18
UR2_POLSP STANDARD; PRT; 12 AA.
ID UR2_POLSP
AC P81022;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE UROTENSIN_II (U-II) (UII).
OS Polyodon spatula (North American paddlefish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Chondrostei; Acipenseriformes; Polyodontidae;
OC Polyodon.
OX NCBI_TaxID=7913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spinal cord;

RX MEDLINE=96051494; PubMed=8536944;
RA Waugh D., Youson J., Mims S.D., Sower S., Conlon J.M.;
RT "Urotensin II from the river lamprey (Lampetra fluviatilis), the sea
lamprey (Petromyzon marinus), and the paddlefish (Polyodon
spatula).";
RL Gen. Comp. Endocrinol. 99:323-332(1995).
DE -1- FUNCTION: HAS A SUGGESTED ROLE IN OSMOREGULATION AND AS A
CORTICOTROPIN-RELEASING FACTOR. PROBABLY INVOLVED IN SMOOTH
MUSCLE STIMULATION.
CC -1- SIMILARITY: BELONGS TO THE UROTENSIN 2 FAMILY.
DR InterPro: IPR001483; Urotensin_II.
DR Pfam: PF02083; Urotensin_II; 1.
DR PROSITE: PS00984; UROTENSIN_II; 1.
KW Hormone.
FT DISULFID 6 11
SQ SEQUENCE 12 AA: 1410 MW: 7551E9DBB879CEBB CRC64;
Query Match 25.0%; Score 2; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 9.8e+02;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 wk 2
DB 8 WK 9
RESULT 19
UR2_SCYCA STANDARD; PRT; 12 AA.
ID UR2_SCYCA
AC P35490;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE UROTENSIN_II (U-II) (UII).
OS Scyliorhinus canicula (Spotted dogfish) (Spotted catshark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphii; Galeoidea; Carcharhiniformes;
OC Scyliorhinidae; Scyliorhinus.
OX NCBI_TaxID=7830;
RN [1]
RP SEQUENCE.
RX TISSUE=Spinal cord;
RX MEDLINE=92319231; PubMed=1620290;
RA Conlon J.M., O'Harte F., Smith D.D., Balmert R.J., Hazen N.;
RT "Purification and characterization of urotensin II and parvalbumin
from an elasmobranch fish, Scyliorhinus canicula (common dogfish).";
RL Neuroendocrinology 55:230-235(1992).
DE -1- FUNCTION: HAS A SUGGESTED ROLE IN OSMOREGULATION AND AS A
CORTICOTROPIN-RELEASING FACTOR. PROBABLY INVOLVED IN SMOOTH
MUSCLE STIMULATION.
CC -1- SIMILARITY: BELONGS TO THE UROTENSIN 2 FAMILY.
DR InterPro: IPR001483; Urotensin_II.
DR Pfam: PF02083; Urotensin_II; 1.
DR PROSITE: PS00984; UROTENSIN_II; 1.
KW Hormone.
FT DISULFID 6 11
SQ SEQUENCE 12 AA: 1526 MW: 804729F9D579CEBA CRC64;
Query Match 25.0%; Score 2; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 9.8e+02;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 wk 2
DB 8 WK 9
RESULT 20
PSBP_PINPS STANDARD; PRT; 13 AA.
ID PSBP_PINPS
AC P81668;

DT 15-JUL-1999 (Rel. 38, Last Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE OXYGEN-EVOLVING ENHANCER PROTEIN 2 (OE2) (23 KDA SUBUNIT OF OXYGEN
DE EVOLVING SYSTEM OF PHOTOSYSTEM II) (FRAGMENT).
GN PSBP.
OS Pinus pinaster (Maritime pine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=71647;
RN (1)
RP TISSUE=Needle;
RC MEDLINE=99274088; PubMed=10344291;
RX Costa P., Plomieu C., Bauw G., Dubos C., Bahrtman N., Kremer A.,
RA Frigerio J.-M., Plomieu C.;
RT "Separation and characterization of needle and xylem maritime pine
RT proteins";
RL Electrophoresis 20:1098-1108(1999).
CC -1- FUNCTION: MAY BE INVOLVED IN THE REGULATION OF PHOTOSYSTEM II.
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST THYLAKOID MEMBRANE; ASSOCIATED
CC WITH THE PHOTOSYSTEM II COMPLEX (BY SIMILARITY).
CC -1- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS PROTEIN
CC (SPOT N179) IS: 5.9. ITS MW IS: 22 KDA.
CC -1- SIMILARITY: BELONGS TO THE PSBP FAMILY.
KM Photosynthesis; Photosystem II; Chloroplast; Thylakoid; Membrane.
FT NON_TER 13
SQ SEQUENCE 13 AA: 1294 MW: C6772B0D54D7C44D CRC64;

Query Match 25.0%; Score 2; DB 1; Length 13;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 5 ay 6
Db 1 AY 2

RESULT 21
ID UVRD_SALTY STANDARD; PRT: 13 AA.
AC 005311;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE DNA HELICASE II (EC 3.6.1.-) (FRAGMENT).
GN UVRD.
OS Salmoneilla typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmoneilla.
OX NCBI_TaxID=602;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=93300795; PubMed=8314774;
RA Smith R.L., Banks J.L., Snavely M.D., Maguire M.E.;
RT "Sequence and topology of the Cora magnesium transport systems of
RT Salmoneilla typhimurium and Escherichia coli. Identification of a new
RT class of transport protein";
RL J. Biol. Chem. 268:14071-14080(1993).
CC -1- FUNCTION: HAS BOTH ATPASE AND HELICASE ACTIVITIES. UNWINDS DNA
CC DUPLEXES WITH 3' TO 5' POLARITY WITH RESPECT TO THE BOUND STRAND
CC AND INITIATES UNWINDING MOST EFFECTIVELY WHEN A SINGLE-STRANDED
CC REGION IS PRESENT. INVOLVED IN THE POSTREPLICATION EVENTS OF
CC NICOTINIC ACETYLCHOLINE REPAIR AND METHYL-DIRECTED MISMATCH REPAIR.
CC -1- SIMILARITY: BELONGS TO THE UVRD SUBFAMILY OF HELICASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.

CC EMBL: L11043; AAA02965.1; -.
DR STYGGEN; SC10414; UVRD.
KM DNA repair; DNA replication; SOS response; Helicase; ATP-D-binding;
KW DNA-binding.
FT NON_TER 1 1
SQ SEQUENCE 13 AA: 1492 MW: D7967B2B8B9ACB5D CRC64;

Query Match 25.0%; Score 2; DB 1; Length 13;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 5 ay 6
Db 6 AY 7

RESULT 22
ID ECDG_LYMDI STANDARD; PRT: 14 AA.
AC P80940;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE TESTIS ECDYSTIOTROPIN PEPTIDE C (TE).
OS Lymantria dispar (Gypsy moth).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Noctuoidea; Lymantriidae; Lymantria.
OX NCBI_TaxID=13123;
RN (1)
RP SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=97387807; PubMed=9243792;
RA Loeb M.U., Wagner R.M., Woods C.W., Gelman D.G., Harrison D.,
RA Bell R.A.;
RT "Naturally occurring analogs of Lymantria testis ecdystiotropin, a
RT gonadotropin isolated from brains of Lymantria dispar pupae.";
RL Arch. Insect Biochem. Physiol. 36:37-50(1997).
CC -1- FUNCTION: STIMULATES SYNTHESIS OF ECDYSTEROID IN THE TESTES
CC OF LARVAE AND PUPAE.
SQ SEQUENCE 14 AA: 1553 MW: 17F479531A685CBB CRC64;

Query Match 25.0%; Score 2; DB 1; Length 14;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 5 ay 6
Db 6 AY 7

RESULT 23
ID MAST_PARID STANDARD; PRT: 14 AA.
AC P42716;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE MASTOPARAN.
OS Parapolybia indica.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
OC Vespoidea; Vespidae; Polistinae; Parapolybia.
OX NCBI_TaxID=31921;
RN (1)
RP SEQUENCE.
RC TISSUE=Venom;
RA Toki T., Yasuhara T., Nakajima T.;
RT "Isolation and sequential analysis of peptides on the venom sac of

RT Parapolybia indica.";
 RL Eisei Dobutsu 39:105-111(1988).
 CC -I- FUNCTION: MAST CELL DEGRANULATING PEPTIDE. ACTIVATES G PROTEINS
 KW THAT COUPLE TO PHOSPHOLIPASE C.
 FT Mast cell degranulation; Venom; Amidation.
 MOD_RES 14
 SEQUENCE 14 AA: 1619 MW: CA376CD3BA6D80DD CRC64;

Query Match 25.0%; Score 2; DB 1; Length 14;
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
 ||
 DB 3 WK 4

RESULT 24
 MAST_POLJA
 ID MAST_POLJA STANDARD; PRT; 14 AA.
 AC P01517;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-AUG-1992 (Rel. 23, Last annotation update)
 DE POLISTES MASTOPARAN.
 OS Polistes jadvigae (Paper wasp).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
 OC Vespididae; Vespidae; Polistinae; Polistes.
 RN NCBI_Taxid=7457;
 RP SEQUENCE.
 RC TISSUE=Venom;
 RA Hirai Y., Ueno Y., Yasuhara T., Yoshida H., Nakajima T.;
 RT "A new mast cell degranulating peptide, polistes mastoparan, in the
 RL venom of Polistes jadvigae.";
 CC Biomed. Res. 1:185-187(1980).
 CC -I- FUNCTION: MAST CELL DEGRANULATING PEPTIDE. ACTIVATES G PROTEINS
 CC THAT COUPLE TO PHOSPHOLIPASE C.
 DR PIR: A01780; QMVAAP.
 KW Mast cell degranulation; Venom; Amidation.
 FT MOD_RES 14
 SEQUENCE 14 AA: 1636 MW: 26472A53BF4778D8 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 14;
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
 ||
 DB 3 WK 4

RESULT 25
 MAST_VESXA
 ID MAST_VESXA STANDARD; PRT; 14 AA.
 AC P01515;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE MASTOPARAN X (MP-X).
 OS Vespa xanthoptera (Japanese hornet).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
 OC Vespoidea; Vespidae; Vespinae; Vespa.
 RN NCBI_Taxid=7448;
 RP SEQUENCE.
 RC TISSUE=Venom;
 RX MEDLINE=80155338; PubMed=540363;
 RA Hirai Y., Kuwada M., Yasuhara T., Yoshida H., Nakajima T.;

RT "A new mast cell degranulating peptide homologous to mastoparan in
 RT the venom of Japanese hornet (Vespa xanthoptera).";
 RL Chem. Pharm. Bull. 27:1945-1946(1979).
 RN [2]
 RP STRUCTURE BY NMR.
 RX MEDLINE=92304942; PubMed=1610813;
 RA Wakamatsu K., Okada A., Miyazawa T., Ohya M., Higashijima T.;
 RT "Membrane-bound conformation of mastoparan-X, a G-protein-activating
 RT peptide.";
 RL Biochemistry 31:5654-5660(1992).
 RN [3]
 RP STRUCTURE BY NMR.
 RX MEDLINE=98206894; PubMed=9537994;
 RA Kusunoki H., Wakamatsu K., Sato K., Miyazawa T., Kohno T.;
 RT "G protein-bound conformation of mastoparan-X: heteronuclear
 RT multidimensional transferred nuclear overhauser effect analysis of
 RT peptide uniformly enriched with ¹³C and ¹⁵N.";
 RL Biochemistry 37:4782-4790(1998).
 CC -I- FUNCTION: MAST CELL DEGRANULATING PEPTIDE. ACTIVATES G PROTEINS
 CC THAT COUPLE TO PHOSPHOLIPASE C.
 DR PIR: A01778; QMVAHX.
 DR PDB: 1A13; 16-FEB-99;
 KW Mast cell degranulation; Venom; Amidation; 3D-structure.
 FT MOD_RES 14
 SEQUENCE 14 AA: 1557 MW: C85DED07AA7AB0DD CRC64;

Query Match 25.0%; Score 2; DB 1; Length 14;
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
 ||
 DB 3 WK 4

Search completed: January 14, 2002, 08:08:21
 Job time: 706 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 08:07:30 ; Search time 81.98 Seconds
(Without alignments)
14.274 Million cell updates/sec

Title: 09-185908-1B
Perfect score: 8
Sequence: 1 wxxxxxxg 8

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues
Word size: 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

Database:

- 1: SP archaea:*
- 2: SP bacteria:*
- 3: SP fungi:*
- 4: SP human:*
- 5: SP invertebrate:*
- 6: SP mammal:*
- 7: SP mhc:*
- 8: SP organelle:*
- 9: SP phage:*
- 10: SP plant:*
- 11: SP rodent:*
- 12: SP virus:*
- 13: SP vertebrate:*
- 14: SP unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	37.5	34	13	057591
2	37.5	100	5	09GY07
3	37.5	163	2	09EM42
4	37.5	171	5	024837
5	37.5	663	8	09MS52
6	25.0	7	8	09S945
7	25.0	8	2	09S443
8	25.0	8	10	P82324
9	25.0	8	11	062721
10	25.0	9	4	09UOM0
11	25.0	9	4	09UMA0
12	25.0	9	5	09TVE1
13	25.0	9	8	09MMG9
14	25.0	9	8	09MMF4
15	25.0	10	8	09XMB4
16	25.0	10	12	086324
17	25.0	10	12	086325
18	25.0	10	12	086326
19	25.0	11	2	P96319

20	25.0	11	2	068237	068237 borrelia bu
21	25.0	11	5	099292	099292 drosophila
22	25.0	11	6	09TRR7	09TRR7 oryctolagus
23	25.0	13	2	055234	055234 synechocyst
24	25.0	13	3	010721	010721 pseudalles
25	25.0	13	4	09UJ52	09UJ52 homo sapien
26	25.0	13	4	09UEE3	09UEE3 homo sapien
27	25.0	13	5	09TWR4	09TWR4 lilyus serr
28	25.0	13	6	09TUV6	09TUV6 ovls arles
29	25.0	13	8	09THR8	09THR8 bryopsis sp
30	25.0	13	10	P82432	P82432 nicotiana t
31	25.0	13	11	P82808	P82808 ratius norv
32	25.0	13	12	P90442	P90442 spodoptera
33	25.0	14	2	P96347	P96347 helicobacte
34	25.0	14	2	09PMT9	09PMT9 campylobact
35	25.0	14	8	09MRU0	09MRU0 arabidopsis
36	25.0	14	9	038469	038469 bacterioph
37	25.0	14	10	P82433	P82433 nicotiana t
38	25.0	14	11	09OVF3	09OVF3 ratius sp.
39	25.0	15	1	09UWH6	09UWH6 thermococu
40	25.0	15	2	053580	053580 rhodobacter
41	25.0	15	2	09R5P2	09R5P2 serratia ma
42	25.0	15	2	09R5I9	09R5I9 bacillus st
43	25.0	15	2	09R5L8	09R5L8 bacillus th
44	25.0	15	2	09R5A1	09R5A1 micrococcu
45	25.0	15	5	09TWL6	09TWL6 echinococu
46	25.0	15	6	09TRG9	09TRG9 bos taurus
47	25.0	15	10	09S820	09S820 hordeum vil
48	25.0	15	11	09OUZ3	09OUZ3 ratius sp.
49	25.0	15	12	069353	069353 herpes slmp
50	25.0	15	12	09PXC1	09PXC1 human immun

ALIGNMENTS

RESULT 1
ID 057591 PRELIMINARY; PRT: 34 AA.
AC 057591;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ARGININOSUCCINATE SYNTHETASE (FRAGMENT).
GN ASS.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Tetraodontiformes; Acanthopterygii; Percomorpha; Tetraodontiformes;
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98077503; PubMed=9414319;
RA Armes N., Gilley J., Fried M.;
RT "The comparative genomic structure and sequence of the surfelt gene
RT homologs in the puffer fish Fugu rubripes and their association with
RT CpG-rich islands.";
RN Genome Res. 7:1138-1152(1997).
[2]
RP SEQUENCE FROM N.A.
RA Fried M.;
RN Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RA Gilley J.;
RT Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: Y15170; CAA75440.1; -.
DR InterPro: IPR001518; Arginosuc_synth.
DR Pfam: PF00764; Arginosuc_synth: 1.
DR ProDom: PD003544; Arginosuc_synth: 1.
FT NON_TER
FT 34
SQ SEQUENCE 34 AA: 3713 MW: 0DC816A13998BA88 CRC64:

Query Match 37.5%; Score 3; DB 13; Length 34;
Best Local Similarity 100.0%; Pred. No. 4.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ayyg 8
||||
DB 9 AXXG 12

RESULT 2

ID 09GX07 PRELIMINARY; PRT; 100 AA.

DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE PROBABLE HYPOTHETICAL 77.1 KD PROTEIN (FRAGMENT).

LM12.160.

OS Leishmania major.

OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.

NCBI_TaxID=5664;

[1]

RP SEQUENCE FROM N.A.

RC STRAIN=FRIDELIN; M. Harris D., Rajandream M., Ivens A., Barrell B.,

RA Oliver K.;

RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

EMBL: AL390114; CAC02042.1; -

DR InterPro: IPR002202; HMG-CoA_red.

DR PROSITE: PS00066; HMG_COA_REDUCTASE_1; UNKNOWN_1.

FT NON_TER 1 100

FT SEQUENCE 100 AA; 11265 MW; 47230C5CD5456206 CRC64;

Query Match 37.5%; Score 3; DB 5; Length 100;

Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wxxxa 5
||||
DB 1 WXXXA 5

RESULT 3

ID 09EM42 PRELIMINARY; PRT; 163 AA.

DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE PENICILLIN BINDING PROTEIN 1A (FRAGMENT).

GN PBP1A.

OS Streptococcus pneumoniae.

OC Bacteria; Firmicutes; Bacilllus/Clostridium group; Streptococcaceae;

NCBI_TaxID=1313;

[1]

RP SEQUENCE FROM N.A.

RC STRAIN=950473;

RA Overweg K., Bogaert D., Sluiter M., de Groot R., Hermans P.W.M.;

RT "Molecular characterization of Streptococcus pneumoniae penicillin-

resistance in the Netherlands."

RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.

EMBL: AJ403979; CAC20961.1; -

FT NON_TER 1 163

FT SEQUENCE 163 AA; 18350 MW; 02B96CAE2DB26ACF CRC64;

Query Match 37.5%; Score 3; DB 2; Length 163;

Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ayyg 8
||||
DB 149 AXXG 152

RESULT 4

ID 024837 PRELIMINARY; PRT; 171 AA.

AC 024837;

DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)

DE OFR 1.5 (FRAGMENT);

OS Entamoeba histolytica.

OC Eukaryota; Entamoebidae; Entamoeba.

NCBI_TaxID=5759;

[1]

RP SEQUENCE FROM N.A.

RC STRAIN=HM-1:JMS;

RL MEDLINE=94099892; PubMed=8274224;

RA Bruchhaus I., Leippe M., Lioutas C., Tannich E.;

RT "Unusual gene organization in the protozoan parasite Entamoeba

histolytica."

RL DNA Cell Biol. 12:925-933(1993).

EMBL: X70851; CAA50201.1; -

FT NON_TER 1 1

FT SEQUENCE 171 AA; 20070 MW; AB7A93347ABA3F74 CRC64;

Query Match 37.5%; Score 3; DB 5; Length 171;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ayyg 8
||||
DB 25 AXXG 28

RESULT 5

ID 09MS52 PRELIMINARY; PRT; 663 AA.

AC 09MS52;

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE NADH DEHYDROGENASE (FRAGMENT).

GN NDHF.

OS Goyazia rupicola.

OC Chloroplast.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; euasterids I; Lamiales; Gesneriaceae; Goyazia.

NCBI_TaxID=125973;

[1]

RP SEQUENCE FROM N.A.

RA Smith J.F.;

RT "The phylogenetic relationships of Lemnocarpos and Goyazia

(Gesneriaceae) based on ndhf sequences."

RL Ann. Mo. Bot. Gard. 0:0-0(2000).

-1- CATALYTIC ACTIVITY: NADH + H⁺ + PLASTOQUINONE = NAD(+) + PLASTOQUINOL.

-1- SIMILARITY: TO NADH-UBIQUINONE/PLASTOQUINONE (COMPLEX I), VARIOUS

CC CHAINS.

EMBL: AF257485; AAF75262.1; -

InterPro: IPR001750; Oxidored_q1.

InterPro: IPR002128; Oxidored_q1_C.

InterPro: IPR001516; Oxidored_q1_N.

Pfam: PF00361; Oxidored_q1; 1.

Pfam: PF01010; Oxidored_q1_C; 1.

Pfam: PF00662; Oxidored_q1_N; 1.

Chloroplast; NAD; Oxidoreductase; Plastocyanine.

FT NON_TER 1 1
NON_TER 663 663
SQ SEQUENCE 663 AA: 74858 MW: BB002498ACE32AE CRC64;

Query Match
Best Local Similarity 37.5%; Score 3; DB 8; Length 663;
Pred. No. 6.8e+03;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ayxg 8
|||
Db 601 AYXG 604

RESULT 6
ID 095945 PRELIMINARY: PRT: 7 AA.
AC 095945:

DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE INSIDE INTRON 5 (FRAGMENT).
OS Saccharomyces cerevisiae (Baker's yeast).
OC Mitochondrion.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;

RP SEQUENCE FROM N.A.
RX MEDLINE=81069885; PubMed=6254986;
RA Bontitz S.G., Coruzzi G., Thalenfeld B., Tagoloff A., Macino G.,
RT "Assembly of the mitochondrial membrane system. Structure and
RT nucleotide sequence of the gene coding for subunit 1 of yeast
RT cytochrome oxidase."
RL J. Biol. Chem. 255:11927-11941(1980).
RL EMBL: V00694; CAA24066.1; -.
KW Mitochondrion.

FT NON_TER 1
SQ SEQUENCE 7 AA: 859 MW: 75B7232362CDC460 CRC64;

Query Match
Best Local Similarity 25.0%; Score 2; DB 8; Length 7;
Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
||
Db 4 WK 5

RESULT 7
ID 09S443 PRELIMINARY: PRT: 8 AA.
AC 09S443:

DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE BETA-LACTAMASE (FRAGMENT).
GN PSE2.

OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;

RP SEQUENCE FROM N.A.
RC STRAIN=R545;

RA Roy D., Coulombe M., Perron K., Roy P.H.;
RT "Characterization of a novel 6'-N-aminoglycoside acetyltransferase
RT gene aac(6')-IIC from the integron of a Chinese Pseudomonas aeruginosa
RT clinical isolate."
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF162771; AAD46628.1; -.
FT NON_TER 8

SQ SEQUENCE 8 AA: 930 MW: EBD85DDDD9D1A336 CRC64;

Query Match
Best Local Similarity 25.0%; Score 2; DB 2; Length 8;
Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6
||
Db 6 AY 7

RESULT 8
ID P82324 PRELIMINARY: PRT: 8 AA.
AC P82324:

DT 01-JUN-2000 (TREMBLrel. 14, Created)
DT 01-JUN-2000 (TREMBLrel. 14, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE UNKNOWN PROTEIN FROM 2D-PAGE OF THYLAKOID (SPOT105) (FRAGMENT).
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Pisum.
OX NCBI_TaxID=3888;

RP SEQUENCE, SUBCELLULAR LOCATION, AND DEVELOPMENTAL STAGE.
RC STRAIN=CV. DE GRACE; TISSUE=LEAF;
RX MEDLINE=20181728; PubMed=10715320;

RA Peltier J.-B., Friso G., Kalume D.E., Koepstorff P., Nilsson F.,
RA Adamaka I., van Wijk R.J.;
RT "Proteomics of the chloroplast: systematic identification and
RT targeting analysis of luminal and peripheral thylakoid proteins."
RL Plant Cell 12:319-341(2000).

CC -1- SUBCELLULAR LOCATION: CHLOROPLAST THYLAKOID MEMBRANE LUMEN OR
CC PERIPHERY.
CC -1- DEVELOPMENTAL STAGE: UNFOLDED AND FULLY DEVELOPED LEAVES.
CC -1- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC PROTEIN IS: 6.7, ITS MW IS: 16.8 KDA.
KW Chloroplast; Thylakoid membrane.

FT NON_TER 1
SQ SEQUENCE 8 AA: 839 MW: DDC68B5DDDC2D2D5 CRC64;

Query Match
Best Local Similarity 25.0%; Score 2; DB 10; Length 8;
Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6
||
Db 5 AY 6

RESULT 9
ID 062721 PRELIMINARY: PRT: 8 AA.
AC 062721:

DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE PROHIBITIN (FRAGMENT).
OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;

RP SEQUENCE FROM N.A.
RC STRAIN=FISHER;
RX MEDLINE=95331633; PubMed=7607556;
RA Altus M.S., Wood C.M., Stewart D.A., Roskams A.I., Friedman V.,
RA Henderson T., Owens G.A., Danner D.B., Jupe E.R., Dell'Orco R.T.,
RA McClung J.K.;

RT "Regions of evolutionary conservation between the rat and human
RT prohibitin-encoding genes."
RL Gene 158:291-294(1995).
DR EMBL; U17178; AAA86692.1; -.
FT NON_TER 8 8
SQ SEQUENCE 8 AA; 1150 MW; EFD3237B05A41376 CRC64;

Query Match 25.0%; Score 2; DB 11; Length 8;
Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
DB 6 wk 7

RESULT 10
O9UOW0 PRELIMINARY; PRT; 9 AA.
AC O9UOW0;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE PROLACTIN PRECURSOR (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=94061059; PubMed=7694728;
RA Spritz R.A., Holmes S.A., Berg S.Z., Nordlund J.J., Fukai K.;
RT "A recurrent deletion in the KIT (mast/stem cell growth factor
RT receptor) proto-oncogene is a frequent cause of human plebaldism."
RL Hum. Mol. Genet. 2:1499-1500(1993).
DR EMBL; S67686; AAD13996.1; -.
FT NON_TER 1 1
SQ SEQUENCE 9 AA; 1182 MW; 0BC504032361B5AB CRC64;

Query Match 25.0%; Score 2; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
DB 5 wk 6

RESULT 12
O9TVF1 PRELIMINARY; PRT; 9 AA.
AC O9TVF1;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE MUCIN-LIKE PROTEIN (FRAGMENT).
GN EMUCE-19C8.
OS Trypanosoma cruzi.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CL-BRENNER;
RX MEDLINE=98225151; PubMed=9556557;
RA Di Noia J.M., D'Orso I., Aslund L., Sanchez D.O., Frasch A.C.;
RT "The Trypanosoma cruzi mucin family is transcribed from hundreds of
RT genes having hypervariable regions."
RL J. Biol. Chem. 273:10843-10850(1998).
DR EMBL; AF036447; AAC14246.1; -.
FT NON_TER 1 1
SQ SEQUENCE 9 AA; 896 MW; DBA831B1B5DD72D CRC64;

Query Match 25.0%; Score 2; DB 5; Length 9;
Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6
DB 4 ay 5

RESULT 13
O9MMG9 PRELIMINARY; PRT; 9 AA.
AC O9MMG9;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT 6 (FRAGMENT).
GN ND6.
OS Buteo buteo vulpinus (western steppe-buzzard).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Falconiformes; Accipitridae;
OC Accipitridae; Buteo.
OX NCBI_TaxID=115228;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VUL-1;
RA Haring E., Riesing M.J., Pinsker W., Gamauf A.;

RP SEQUENCE FROM N.A.
RX MEDLINE=94061059; PubMed=7694728;
RA Spritz R.A., Holmes S.A., Berg S.Z., Nordlund J.J., Fukai K.;
RT "A recurrent deletion in the KIT (mast/stem cell growth factor
RT receptor) proto-oncogene is a frequent cause of human plebaldism."
RL Hum. Mol. Genet. 2:1499-1500(1993).
DR EMBL; S67686; AAD13996.1; -.
FT NON_TER 1 1
SQ SEQUENCE 9 AA; 1182 MW; 0BC504032361B5AB CRC64;

Query Match 25.0%; Score 2; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
DB 5 wk 6

RESULT 12
O9TVF1 PRELIMINARY; PRT; 9 AA.
AC O9TVF1;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE MUCIN-LIKE PROTEIN (FRAGMENT).
GN EMUCE-19C8.
OS Trypanosoma cruzi.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CL-BRENNER;
RX MEDLINE=98225151; PubMed=9556557;
RA Di Noia J.M., D'Orso I., Aslund L., Sanchez D.O., Frasch A.C.;
RT "The Trypanosoma cruzi mucin family is transcribed from hundreds of
RT genes having hypervariable regions."
RL J. Biol. Chem. 273:10843-10850(1998).
DR EMBL; AF036447; AAC14246.1; -.
FT NON_TER 1 1
SQ SEQUENCE 9 AA; 896 MW; DBA831B1B5DD72D CRC64;

Query Match 25.0%; Score 2; DB 5; Length 9;
Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6
DB 4 ay 5

RESULT 13
O9MMG9 PRELIMINARY; PRT; 9 AA.
AC O9MMG9;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT 6 (FRAGMENT).
GN ND6.
OS Buteo buteo vulpinus (western steppe-buzzard).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Falconiformes; Accipitridae;
OC Accipitridae; Buteo.
OX NCBI_TaxID=115228;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VUL-1;
RA Haring E., Riesing M.J., Pinsker W., Gamauf A.;

```
RT "Evolution of a pseudo-control region in the mitochondrial genome of
RT Palearctic Buzzards (genus Buteo).";
RL J. Zool. Syst. Evol. Res. 37:185-194(1999).
DR EMBL: AF202197; AAF61879.1; -.
KW Mitochondrion.
FT NON_TER
SQ SEQUENCE 9 AA: 1026 MW: C24E272732C9DB5D CRC64:

Query Match 25.0%; Score 2; DB 8; Length 9;
Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ay 6
DB 2 AY 3

RESULT 14
Q9MWF4 PRELIMINARY; PRT; 9 AA.
ID Q9MWF4;
AC Q9MWF4;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DR MADH DEHYDROGENASE SUBUNIT 6 (FRAGMENT).
GN ND6.
OS Buteo rufinus rufinus (Long-legged buzzard).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Falconiformes; Accipitridae;
OC Accipitrinae; Buteo.
OX NCBI_TaxID=116585;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RUF-1;
RA Haring E., Rising M.J., Pinsker W., Gamauf A.;
RT "Evolution of a pseudo-control region in the mitochondrial genome of
RT Palearctic Buzzards (genus Buteo).";
RL J. Zool. Syst. Evol. Res. 37:185-194(1999).
DR EMBL: AF20212; AAF61894.1; -.
KW Mitochondrion.
FT NON_TER
SQ SEQUENCE 9 AA: 1026 MW: C24E272732C9DB5D CRC64:

Query Match 25.0%; Score 2; DB 8; Length 9;
Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ay 6
DB 2 AY 3

RESULT 15
Q9XMB4 PRELIMINARY; PRT; 10 AA.
ID Q9XMB4;
AC Q9XMB4;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE MADH DEHYDROGENASE SUBUNIT 3 (FRAGMENT).
GN MAD3.
OS Aegilops tauschii (Patropyrum tauschii).
OC Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticaceae; Aegilops.
OX NCBI_TaxID=37682;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KU29;
```

```
RA Tsukamoto N., Asakura N., Takumi S., Mori N., Nakamura C.;
RT "The presence of paternal sub-genomic mitochondrial DNA copies in the
RT nucleus-cytoplasm hybrids of tetraploid wheat with the cytoplasm of
RT Aegilops squarrosa.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF142479; AAD37355.1; -.
KW Mitochondrion.
FT NON_TER
SQ SEQUENCE 10 AA: 1233 MW: 5F9A1B5BD86403 CRC64:

Query Match 25.0%; Score 2; DB 8; Length 10;
Best Local Similarity 100.0%; Pred. No. 3.6e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wk 2
DB 2 WK 3

RESULT 16
Q86324 PRELIMINARY; PRT; 10 AA.
ID Q86324;
AC Q86324;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1996 (TReMBLrel. 01, Last annotation update)
DE GP37 (FRAGMENT).
GN ENV.
OS Rous sarcoma virus.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11886;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SCHMIDT-RUPPIN SUBGROUP A (NY);
RA Hara H., Kaji A.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: U41726; AAB60580.1; -.
FT NON_TER
SQ SEQUENCE 10 AA: 1119 MW: 27ED4115BB0776D8 CRC64:

Query Match 25.0%; Score 2; DB 12; Length 10;
Best Local Similarity 100.0%; Pred. No. 3.6e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ay 6
DB 1 AY 2

RESULT 17
Q86325 PRELIMINARY; PRT; 10 AA.
ID Q86325;
AC Q86325;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1996 (TReMBLrel. 01, Last annotation update)
DE GP37 (FRAGMENT).
GN ENV.
OS Rous sarcoma virus.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11886;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SCHMIDT-RUPPIN SUBGROUP A (NY);
RA Hara H., Kaji A.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: U41727; AAB60581.1; -.
FT NON_TER
SQ SEQUENCE 10 AA: 1119 MW: 27ED4115BB0776D8 CRC64;
```

Query Match 25.0%; Score 2; DB 12; Length 10;
Best Local Similarity 100.0%; Pred. No. 3.6e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6
11
Db 1 AY 2

RESULT 18
O68326 PRELIMINARY; PRT; 10 AA.
ID O68326;
AC O68326;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE GP37 (FRAGMENT).
GN ENV.
OS Rous sarcoma virus.
OC Viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11886;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SCHMIDT-RUPPIN SUBGROUP A (NY);
RA Hara H., Kaji A.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SCHMIDT-RUPPIN SUBGROUP A (NY);
RX MEDLINE=82271824; PubMed=6287213;
RA Takeya T., Hanafusa H., Junghans R.P., Ju G., Skalka A.M.;
RT "Comparison between the viral transforming gene (src) of recovered
avian sarcoma virus and its cellular homolog.";
RL Mol. Cell. Biol. 1:1024-1037(1981).
DR EMBL: U41729; AAA84421.1; -.
FT NON_TER 1
SQ SEQUENCE 10 AA; 1119 MW; 27ED415B80776D8 CRC64;

Query Match 25.0%; Score 2; DB 12; Length 10;
Best Local Similarity 100.0%; Pred. No. 3.6e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6
11
Db 1 AY 2

RESULT 19
P96319 PRELIMINARY; PRT; 11 AA.
ID P96319;
AC P96319;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE CODD PORTION OF PROTEOLYSIS TAG (FRAGMENT).
OS Desulfovibrio desulfuricans.
OC Bacteria; Proteobacteria; delta subdivision; Desulfovibrio.
OX NCBI_TaxID=876;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 27774;
RX MEDLINE=97128184; PubMed=8972778;
RA Williams K.P., Bartel D.P.;
RT "Phylogenetic analysis of tmRNA secondary structure.";
DR EMBL: U68081; AAA848023.1; -.
FT NON_TER 1
SQ SEQUENCE 11 AA; 1250 MW; 85776D58CB5AB5A CRC64;

Query Match 25.0%; Score 2; DB 2; Length 11;
Best Local Similarity 100.0%; Pred. No. 3.9e+03;

Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6
11
Db 8 AY 9

RESULT 20
O68237 PRELIMINARY; PRT; 11 AA.
ID O68237;
AC O68237;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
DE PLASMID CP32-4, POSSIBLE PARTITION PROTEINS, COMPLETE CDS (FRAGMENT).
OS Borrelia burgdorferi (Lyme disease spirochete).
OC plasmid cp32-4.
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B31;
RA Stevenson B., Casjens S., Rosa P.;
RL Microbiology 0:0-0(1998).
DR EMBL: AF022481; AAC35449.1; -.
KW plasmid.
FT NON_TER 11
SQ SEQUENCE 11 AA; 1237 MW; 50E3B714D45B5D7 CRC64;

Query Match 25.0%; Score 2; DB 2; Length 11;
Best Local Similarity 100.0%; Pred. No. 3.9e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6
11
Db 5 AY 6

RESULT 21
Q99292 PRELIMINARY; PRT; 11 AA.
ID Q99292;
AC Q99292;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE BICOID PROTEIN (FRAGMENT).
GN BCD.
OS Drosophila heteroneura (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=32382;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91184004; PubMed=2081457;
RA MacDonald P.M.;
RT "bicoid mRNA localization signal: phylogenetic conservation of
function and RNA secondary structure.";
KL Development 110:161-171(1990).
CC -I- FUNCTION: BICOID IS SEGMENT-POLARITY PROTEIN THAT PROVIDES
CC POSTERIORAL CUES FOR THE DEVELOPMENT OF HEAD AND THORACIC SEGMENTS.
CC BCD REGULATES THE EXPRESSION OF ZYGOTIC GENES, POSSIBLY THROUGH
CC ITS HOMODOMAIN, AND INHIBITS THE ACTIVITY OF OTHER MATERNAL GENE
CC PRODUCTS. IT IS POSSIBLE THAT BCD ALSO BINDS RNA.
DR EMBL: M32125; AAA28386.1; -.
DR Flybase: FBgn012352; Dmel\bcd.
RW Homobox; DNA-binding; Developmental protein; Nuclear protein;
KW Segmentation polarity protein; Transcription regulation; RNA-binding.
FT NON_TER 1
SQ SEQUENCE 11 AA; 1221 MW; 8CE802305D9D6C1 CRC64;

Query Match 25.0%; Score 2; DB 5; Length 11;
Best Local Similarity 100.0%; Pred. No. 3.9e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0;

OY 5 ay 6
DB 7 AY 8

RESULT 22
O9TRR7 PRELIMINARY; PRT; 11 AA.
AC O9TRR7:
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE CALCYCICLIN-ASSOCIATED PROTEIN, CAP50=CA2+/PHOSPHOLIPID-BINDING PROTEIN L-13 FRAGMENT.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE.
RX MEDLINE=92250478; PubMed=1533622;
RA Tokumitsu H., Mizutani A., Minami H., Kobayashi R., Hidaka H.;
"A calcyciclin-associated protein is a newly identified member of the Ca2+/phospholipid-binding proteins, annexin family."
RL J. Biol. Chem. 267:8919-8924(1992).
SO SEQUENCE 11 AA; 1310 MW; 55580B0F5DDAA9C7 CRC64;

Query Match 25.0%; Score 2; DB 6; Length 11;
Best Local Similarity 100.0%; Pred. No. 3.9e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0;

OY 5 ay 6
DB 7 AY 8

RESULT 23
O55234 PRELIMINARY; PRT; 13 AA.
AC O55234:
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)
DE 2-HYDROXYACID DEHYDROGENASE HOMOLOGUE (FRAGMENT).
OS Synecocystis sp. (Strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC6803;
RA Terauchi K., Ikeuchi M., Ohmori M.;
"A putative Fd-GOGAT gene involved in protection against RT photoinhibition in Synecocystis PCC 6803."
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: D78371; BAA11378.1; -.
FT NON_TER 13 13
SO SEQUENCE 13 AA; 1564 MW; CC84E4282B1CA5B9 CRC64;

Query Match 25.0%; Score 2; DB 2; Length 13;
Best Local Similarity 100.0%; Pred. No. 4.6e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0;

OY 5 ay 6
DB 10 AY 11

RESULT 24
Q10721 PRELIMINARY; PRT; 13 AA.
AC Q10721:
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 06, Last annotation update)
DE ALKALINE EXTRACELLULAR PROTEASE (EC 3.4.21.-) (AEP) (FRAGMENT).
OS Pseudallescheria boydii.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Microscopales; Microascales; Pseudallescheria.
OX NCBI_TaxID=5597;
RN [1]
RP SEQUENCE.
RX MEDLINE=96207571; PubMed=8670095;
RA Larcher G., Cimon B., Symons F., Tronchin G., Chabasse D.,
Bouchara J.-P.;
"A 33 kDa serine proteinase from Scedosporium apiospermum."
RL Biochem. J. 315:119-126(1996).
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE CC SUBTILASE FAMILY.
KM Hydrolase; Serine protease; Zymogen.
FT NON_TER 13 13
SO SEQUENCE 13 AA; 1292 MW; 9DD5F3294A68D861 CRC64;

Query Match 25.0%; Score 2; DB 3; Length 13;
Best Local Similarity 100.0%; Pred. No. 4.6e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0;

OY 5 ay 6
DB 1 AY 2

RESULT 25
Q90J52 PRELIMINARY; PRT; 13 AA.
AC Q90J52:
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE LATROPHILIN-2 (FRAGMENT).
GN LPHN1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99153747; PubMed=10030676;
RA White G.R.M., Varley J.M., Helgway J.;
"Isolation and characterisation of a human homologue of the RT latrophilin gene from a region of 1p31.1 implicated in breast cancer."
RL Oncogene 17:3513-3519(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA White G.R.M., Varley J.M., Helgway J.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ244500; CAB60202.1; -.
FT NON_TER 1 1
FT NON_TER 13 13
SO SEQUENCE 13 AA; 1637 MW; 4161F9BECF72A1A3 CRC64;

Query Match 25.0%; Score 2; DB 4; Length 13;
Best Local Similarity 100.0%; Pred. No. 4.6e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0;

OY 5 ay 6

Db 11
10 AY 11

Search completed: January 14, 2002, 08:07:31
Job time: 761 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:56:28 : Search time 81.39 Seconds
(without alignments)
7.281 Million cell updates/sec

Title: 09-185908-1b
Perfect score: 8
Sequence: 1 wxxxxxxg 8

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: listing first 50 summaries

Database :

A.Geneseq_1101:*

1: /SID52/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq/AA2001.DAT:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3	37.5	9	21	AAV51341	Sphingolipid desat
2	3	37.5	13	22	AAW37631	Cholecystokinin pe
3	3	37.5	91	21	AAW22631	Zea mays protein f
4	3	37.5	110	18	AAW27779	Amino acid sequenc
5	3	37.5	121	21	AAW22530	Zea mays protein f
6	3	37.5	185	18	AAW20262	H. pylori secreted
7	3	37.5	185	18	AAW24622	H. pylori secreted
8	3	37.5	275	12	AAW10204	Mutant subtilisin
9	3	37.5	277	22	AAW75013	Human colon cancer
10	3	37.5	284	21	AAW56744	Human prostate can
11	3	37.5	401	12	AAW12129	ORF 1 of IgG light

12	3	37.5	489	20	AAW84298	Consensus sequence
13	3	37.5	498	20	AAW84183	Consensus sequence
14	3	37.5	500	22	AAW40758	Human polypeptide
15	3	37.5	513	20	AAW38787	Neisseria meningit
16	3	37.5	515	9	AAW80575	Mutated alpha-amy1
17	3	37.5	688	17	AAW17569	Bacillus CGTase va
18	3	37.5	688	17	AAW17577	Bacillus CGTase va
19	3	37.5	688	17	AAW17573	Bacillus CGTase va
20	3	37.5	689	17	AAW17570	Bacillus CGTase va
21	3	37.5	689	17	AAW17574	Bacillus CGTase va
22	3	37.5	689	17	AAW17578	Bacillus CGTase va
23	2	25.0	3	15	AAW42556	ACE inhibitor Sp3.
24	2	25.0	4	2	AAW10370	Generic enkephalin
25	2	25.0	4	2	AAW10372	Enkephalin-like an
26	2	25.0	4	2	AAW10373	Enkephalin-like an
27	2	25.0	4	2	AAW10375	Enkephalin-like an
28	2	25.0	4	2	AAW10386	Generic analgesic
29	2	25.0	4	2	AAW10599	N-adamantane tetra
30	2	25.0	4	2	AAW10433	Analgesic tetrapep
31	2	25.0	4	2	AAW10401	Agonist peptide.
32	2	25.0	4	2	AAW10620	Analgesic tetrapep
33	2	25.0	4	2	AAW10625	Analgesic tetrapep
34	2	25.0	4	3	AAW20208	Analgesic and neur
35	2	25.0	4	3	AAW20210	Analgesic and neur
36	2	25.0	4	5	AAW40339	Sequence of enkeph
37	2	25.0	4	6	AAW50545	Sequence of cyclo
38	2	25.0	4	7	AAW61656	Sequence of peptid
39	2	25.0	4	8	AAW71313	Peptide component
40	2	25.0	4	9	AAW81591	Organ specific neo
41	2	25.0	4	11	AAW04767	Cyclic enkephalin
42	2	25.0	4	11	AAW07020	Enkephalin analogu
43	2	25.0	4	12	AAW30332	Iserin deriv. syn
44	2	25.0	4	12	AAW11225	Ketone analogue pr
45	2	25.0	4	12	AAW11228	Ketone analogue pr
46	2	25.0	4	12	AAW12507	Enkephalin analogu
47	2	25.0	4	12	AAW12508	5 amino acid resid
48	2	25.0	4	13	AAW20049	Tetrapeptide chole
49	2	25.0	4	15	AAW42585	Kyotorphin analogu
50	2	25.0	4	15	AAW42588	Kyotorphin analogu

ALIGNMENTS

RESULT 1	
AAW51341	
ID	AAW51341 standard; Protein: 9 AA.
AC	AAW51341:
XX	
DT	27-APR-2000 (first entry)
XX	
DE	Sphingolipid desaturase protein fragment #7.
XX	
KW	Sphingolipid desaturase; sld1; sphingobase; ceramide; capnoid;
KW	transgenic plant; crop plant; delta-8-unsaturated long-chain base;
KW	tolerance; resistance; soil salinity; ion stress; toxicity; drought;
KW	cold; frost; phytopathogenic microorganism; flowering time; cosmetic;
KW	pharmaceutical; food; chemical raw material.
XX	
OS	Unidentified.
XX	
FH	Key
FT	Misc-difference 3
XX	Location/Qualifiers
XX	Any-amino_acid
PN	DE19828850-A1.
XX	
PD	30-DEC-1999.
XX	
PF	27-JUN-1998; 98DE-1028850.
XX	
PR	27-JUN-1998; 98DE-1028850.

XX (GVSE-) GVS GES ERWERB & VERW LANDWIRTSCHAFTLICH.
XX Heinz E, Zaehrerger U, Schmidt H, Sperling P;
PI WPI: 2000-127549/12.
XX
XX New sphingolipid desaturase that selectively introduces double bond
PT into sphingolipids and capnoids -
PS
PS Claim 4; Page 23; 62pp; German.
XX
XX This invention describes a novel sphingolipid desaturase that
CC selectively introduces a double bond into the sphingobase of the ceramide
CC residue of sphingolipids and capnoids. A DNA sequence encoding the
CC sphingolipid desaturase, or a vector containing the DNA sequence, can be
CC used to produce transgenic plants, especially crop plants, with an
CC increased or decreased delta-8-unsaturated long-chain base content or an
CC altered delta-8-unsaturated long-chain base cis/trans ratio, especially
CC to compensate for a delta-8-unsaturated long-chain base deficiency, to
CC exclude production of delta-8-unsaturated bases, to increase tolerance
CC or resistance to soil salinity, ion stress or toxicity, drought, wet
CC conditions, cold or frost and/or phytopathogenic microorganisms, or to
CC alter size growth and flowering time. Cells, transgenic organisms or
CC plants containing the DNA sequence can be used to produce sphingolipids
CC and capnoids with unsaturated sphingobases. The sphingolipids or
CC capnoids can be used in cosmetics, pharmaceuticals and foods and as
CC chemical raw materials. AA51335-Y51344 represent sphingolipid desaturase
CC protein fragments described in the method of the invention.
XX
SQ Sequence 9 AA;

Query Match 37.5%; Score 3; DB 21; Length 9;
Best Local Similarity 100.0%; Pred. No. 4.3e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 ayxg 8
|||
Db 1 ayxg 4

RESULT 2
AAB37631
ID AAB37631 standard; peptide; 13 AA.
XX
AC AAB37631;
XX
DT 23-MAR-2001 (first entry)
XX
DE Cholecystokinin peptide fragment analogue #10.
XX
KW Antidiabetic; cytostatic; auditory; cholecystokinin-8; CCK-8; neuropathy;
KW peripheral nervous system; diabetes mellitus; cancer treatment;
KW cytostatic; hearing impairment; visual handicap;
KW alcohol-induced neuropathy; dystrophy.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 8
FT Modified-site /label= Nle
FT Modified-site /label= Nle
FT Modified-site /label= Nle
FT Modified-site /note= "C-terminal amide"
XX
XX WO200066150-A1.
XX
PD 09-NOV-2000.
XX
XX 03-MAY-2000; 2000WO-SE00870.
PF
XX

PR 03-MAY-1999; 99SE-0001578.
XX
XX (KARO-) KAROLINSKA INNOVATIONS AB.
XX
XX Lundeborg T, Manni L;
XX
XX WPI: 2001-024739/03.
XX
XX Manufacturing a medicament for treating neuropathies in the peripheral
PT nervous system comprises use of a substance showing cholecystokinin-8
PT activity -
PS
PS Disclosure; Page 4; 38pp; English.
XX
XX The present invention relates to peptides showing cholecystokinin (CCK)-8
CC activity, which can be used to treat neuropathies in the peripheral
CC nervous system (PNS). The present sequence is one such peptide. The
CC peptides of the present invention may be used to treat neuropathies in
CC the PNS associated with diabetes mellitus, cancer treatment such as
CC cytostatic, hearing impairment and/or visual handicap, alcohol-induced
CC neuropathy, damage induced by surgery and dystrophy.
XX
SQ Sequence 13 AA;

Query Match 37.5%; Score 3; DB 22; Length 13;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 ayxg 8
|||
Db 6 ayxg 9

RESULT 3
AAG22631
ID AAG22631 standard; Protein; 91 AA.
XX
AC AAG22631;
XX
DT 17-OCT-2000 (first entry)
XX
DE Zea mays protein fragment SEQ ID NO: 25635.
XX
DE Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.
XX
OS Zea mays subsp. mays.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999;	99US-0132484.	PR 22-JUL-1999;	99US-0145085.
PR 05-MAY-1999;	99US-0132485.	PR 22-JUL-1999;	99US-0145087.
PR 06-MAY-1999;	99US-0132486.	PR 22-JUL-1999;	99US-0145089.
PR 06-MAY-1999;	99US-0132487.	PR 22-JUL-1999;	99US-0145192.
PR 07-MAY-1999;	99US-0132863.	PR 23-JUL-1999;	99US-0145145.
PR 11-MAY-1999;	99US-0134256.	PR 23-JUL-1999;	99US-0145218.
PR 14-MAY-1999;	99US-0134218.	PR 23-JUL-1999;	99US-0145224.
PR 14-MAY-1999;	99US-0134219.	PR 26-JUL-1999;	99US-0145276.
PR 14-MAY-1999;	99US-0134321.	PR 27-JUL-1999;	99US-0145913.
PR 14-MAY-1999;	99US-0134370.	PR 27-JUL-1999;	99US-0145918.
PR 18-MAY-1999;	99US-0134768.	PR 27-JUL-1999;	99US-0145919.
PR 19-MAY-1999;	99US-0134941.	PR 28-JUL-1999;	99US-0145951.
PR 20-MAY-1999;	99US-0135124.	PR 02-AUG-1999;	99US-0146386.
PR 21-MAY-1999;	99US-0135353.	PR 02-AUG-1999;	99US-0146388.
PR 24-MAY-1999;	99US-0135629.	PR 02-AUG-1999;	99US-0146388.
PR 25-MAY-1999;	99US-0136021.	PR 03-AUG-1999;	99US-0147303.
PR 27-MAY-1999;	99US-0136392.	PR 03-AUG-1999;	99US-0147204.
PR 28-MAY-1999;	99US-0136782.	PR 04-AUG-1999;	99US-0147302.
PR 01-JUN-1999;	99US-0137222.	PR 05-AUG-1999;	99US-0147192.
PR 03-JUN-1999;	99US-0137528.	PR 05-AUG-1999;	99US-0147260.
PR 04-JUN-1999;	99US-0137502.	PR 06-AUG-1999;	99US-0147303.
PR 07-JUN-1999;	99US-0137724.	PR 06-AUG-1999;	99US-0147416.
PR 08-JUN-1999;	99US-0138094.	PR 09-AUG-1999;	99US-0147493.
PR 10-JUN-1999;	99US-0138540.	PR 09-AUG-1999;	99US-0147935.
PR 10-JUN-1999;	99US-0138847.	PR 10-AUG-1999;	99US-0148171.
PR 14-JUN-1999;	99US-0139119.	PR 11-AUG-1999;	99US-0148319.
PR 16-JUN-1999;	99US-0139452.	PR 12-AUG-1999;	99US-0148341.
PR 16-JUN-1999;	99US-0139453.	PR 13-AUG-1999;	99US-0148565.
PR 17-JUN-1999;	99US-0139492.	PR 13-AUG-1999;	99US-0148684.
PR 18-JUN-1999;	99US-0139454.	PR 16-AUG-1999;	99US-0149368.
PR 18-JUN-1999;	99US-0139455.	PR 17-AUG-1999;	99US-0149173.
PR 18-JUN-1999;	99US-0139456.	PR 18-AUG-1999;	99US-0149173.
PR 18-JUN-1999;	99US-0139457.	PR 18-AUG-1999;	99US-0149426.
PR 18-JUN-1999;	99US-0139458.	PR 20-AUG-1999;	99US-0149722.
PR 18-JUN-1999;	99US-0139459.	PR 20-AUG-1999;	99US-0149723.
PR 18-JUN-1999;	99US-0139460.	PR 20-AUG-1999;	99US-0149929.
PR 18-JUN-1999;	99US-0139461.	PR 23-AUG-1999;	99US-0149902.
PR 18-JUN-1999;	99US-0139462.	PR 23-AUG-1999;	99US-0149930.
PR 18-JUN-1999;	99US-0139463.	PR 25-AUG-1999;	99US-0150566.
PR 18-JUN-1999;	99US-0139463.	PR 26-AUG-1999;	99US-0150884.
PR 18-JUN-1999;	99US-0139750.	PR 27-AUG-1999;	99US-0151065.
PR 18-JUN-1999;	99US-0139763.	PR 27-AUG-1999;	99US-0151066.
PR 21-JUN-1999;	99US-0139817.	PR 27-AUG-1999;	99US-0151080.
PR 22-JUN-1999;	99US-0139899.	PR 30-AUG-1999;	99US-0151303.
PR 23-JUN-1999;	99US-0140353.	PR 31-AUG-1999;	99US-0151438.
PR 23-JUN-1999;	99US-0140354.	PR 01-SEP-1999;	99US-0152363.
PR 24-JUN-1999;	99US-0140695.	PR 07-SEP-1999;	99US-0152363.
PR 28-JUN-1999;	99US-0140823.	PR 10-SEP-1999;	99US-0153070.
PR 29-JUN-1999;	99US-0140991.	PR 13-SEP-1999;	99US-0153758.
PR 30-JUN-1999;	99US-0141287.	PR 15-SEP-1999;	99US-0154018.
PR 01-JUL-1999;	99US-0141842.	PR 16-SEP-1999;	99US-0154039.
PR 01-JUL-1999;	99US-0142154.	PR 20-SEP-1999;	99US-0154779.
PR 02-JUL-1999;	99US-0142055.	PR 22-SEP-1999;	99US-0155139.
PR 06-JUL-1999;	99US-0142390.	PR 23-SEP-1999;	99US-0155486.
PR 08-JUL-1999;	99US-0142803.	PR 24-SEP-1999;	99US-0155659.
PR 09-JUL-1999;	99US-0142920.	PR 28-SEP-1999;	99US-0156458.
PR 12-JUL-1999;	99US-0142977.	PR 29-SEP-1999;	99US-0156596.
PR 13-JUL-1999;	99US-0143542.	PR 04-OCT-1999;	99US-0157117.
PR 14-JUL-1999;	99US-0143624.	PR 05-OCT-1999;	99US-0157753.
PR 15-JUL-1999;	99US-0144005.	PR 06-OCT-1999;	99US-0157865.
PR 16-JUL-1999;	99US-0144085.	PR 07-OCT-1999;	99US-0158029.
PR 16-JUL-1999;	99US-0144086.	PR 08-OCT-1999;	99US-0158232.
PR 19-JUL-1999;	99US-0144325.	PR 12-OCT-1999;	99US-0158369.
PR 19-JUL-1999;	99US-0144331.	PR 13-OCT-1999;	99US-0159293.
PR 19-JUL-1999;	99US-0144332.	PR 13-OCT-1999;	99US-0159294.
PR 19-JUL-1999;	99US-0144333.	PR 13-OCT-1999;	99US-0159295.
PR 19-JUL-1999;	99US-0144334.	PR 14-OCT-1999;	99US-0159329.
PR 19-JUL-1999;	99US-0144335.	PR 14-OCT-1999;	99US-0159330.
PR 20-JUL-1999;	99US-0144332.	PR 14-OCT-1999;	99US-0159637.
PR 20-JUL-1999;	99US-0144884.	PR 14-OCT-1999;	99US-0159638.
PR 21-JUL-1999;	99US-0144814.	PR 18-OCT-1999;	99US-0159584.
PR 21-JUL-1999;	99US-0145086.	PR 21-OCT-1999;	99US-0160741.
PR 21-JUL-1999;	99US-0145088.	PR 21-OCT-1999;	99US-0160767.

PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 37.5%; Score 3; DB 21; Length 91;
Best Local Similarity 100.0%; Pred. No. 5.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ayxg 8
|||
Db 27 ayxg 30

RESULT 4

AAW27779
ID AAW27779 standard; Protein: 110 AA.

AC AAW27779;

DT 21-JUL-1998 (first entry)

XX Amino acid sequence of pseudouridylylate synthase I.

XX Staphylococcus aureus protein; ribozyme; antisense sequence; control;

KW Staphylococcal gene; regulatory element; bacterial gene expression;

KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;

KW toxic shock syndrome.

XX Staphylococcus aureus.

XX Key Location/Qualifiers

FT Misc-difference 10 /note= "not specified"

FT Misc-difference 18 /note= "not specified"

PN WO9730070-A1.

PD 21-AUG-1997.

XX 19-FEB-1997; 97WO-US02318.

XX 20-FEB-1996; 96US-0011888.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO;

PI Pratt JM, Reichard RW, Rosenberg M, Ward JM,

XX WPI: 1997-424969/39.

DR N-PSDB: MAT83748.

XX Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used

PT to isolate antimicrobial compounds, and in vaccines against S.

PT aureus infection

XX Claim 6; Page 270; 989pp; English.

XX The present sequence represents a Staphylococcus aureus protein, that,

CC based on homology with an Escherichia coli protein, is believed
CC to be pseudouridylylate synthase I (pseudouridine synthase I, uracil
CC hydrolyase. The DNA sequence was isolated from a library of clones of
CC S. aureus WCUH 29 in Escherichia coli. The DNA sequence can be used in
CC the construction of ribozymes and antisense sequences to control the
CC expression of Staphylococcal genes. The DNA sequence is also useful as
CC a source of regulatory elements for the control of bacterial gene
CC expression. The present protein may be used to produce vaccines to
CC enable a host to produce specific antibodies with antibacterial action.
CC These vaccines and antibodies would protect a host against invasion by
CC S. aureus, and conditions relating to Staphylococcal infection,
CC e.g. Staphylococcal food poisoning, scaled skin syndrome, and toxic
CC shock syndrome.

Query Match 37.5%; Score 3; DB 18; Length 110;
Best Local Similarity 100.0%; Pred. No. 6.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ayxg 8
|||
Db 8 ayxg 11

RESULT 5

AAAG22630
ID AAAG22630 standard; Protein: 121 AA.

AC AAAG22630;

DT 17-OCT-2000 (first entry)

XX Zea mays protein fragment SEQ ID NO: 25634.

XX Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence; corn.

XX Zea mays subsp. mays.

XX EP1033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 23-APR-1999; 99US-0130891.

XX 28-APR-1999; 99US-0131449.

XX 30-APR-1999; 99US-0132048.

XX 30-APR-1999; 99US-0132407.

XX 04-MAY-1999; 99US-0132484.

XX 05-MAY-1999; 99US-0132485.

XX 06-MAY-1999; 99US-0132486.

XX 07-MAY-1999; 99US-0132487.

XX 11-MAY-1999; 99US-0134256.

XX 14-MAY-1999; 99US-0134218.

XX 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999;	99US-0134221.	PR 27-JUL-1999;	99US-0145913.
PR 14-MAY-1999;	99US-0134570.	PR 27-JUL-1999;	99US-0145918.
PR 18-MAY-1999;	99US-0134768.	PR 27-JUL-1999;	99US-0145919.
PR 19-MAY-1999;	99US-0134841.	PR 28-JUL-1999;	99US-0145951.
PR 20-MAY-1999;	99US-0135124.	PR 02-AUG-1999;	99US-0146386.
PR 21-MAY-1999;	99US-0135353.	PR 02-AUG-1999;	99US-0146388.
PR 24-MAY-1999;	99US-0136229.	PR 02-AUG-1999;	99US-0146389.
PR 25-MAY-1999;	99US-0136021.	PR 03-AUG-1999;	99US-0147038.
PR 27-MAY-1999;	99US-0136392.	PR 04-AUG-1999;	99US-0147204.
PR 28-MAY-1999;	99US-0136782.	PR 04-AUG-1999;	99US-0147302.
PR 01-JUN-1999;	99US-0137322.	PR 05-AUG-1999;	99US-0147192.
PR 03-JUN-1999;	99US-0137528.	PR 05-AUG-1999;	99US-0147260.
PR 04-JUN-1999;	99US-0137502.	PR 06-AUG-1999;	99US-0147303.
PR 07-JUN-1999;	99US-0137724.	PR 06-AUG-1999;	99US-0147416.
PR 08-JUN-1999;	99US-0138094.	PR 09-AUG-1999;	99US-0147493.
PR 10-JUN-1999;	99US-0138540.	PR 09-AUG-1999;	99US-0147935.
PR 10-JUN-1999;	99US-0138847.	PR 10-AUG-1999;	99US-0148171.
PR 14-JUN-1999;	99US-0139119.	PR 11-AUG-1999;	99US-0148319.
PR 16-JUN-1999;	99US-0139452.	PR 12-AUG-1999;	99US-0148341.
PR 16-JUN-1999;	99US-0139453.	PR 13-AUG-1999;	99US-0148565.
PR 17-JUN-1999;	99US-0139492.	PR 13-AUG-1999;	99US-0148684.
PR 18-JUN-1999;	99US-0139454.	PR 16-AUG-1999;	99US-0149368.
PR 18-JUN-1999;	99US-0139455.	PR 17-AUG-1999;	99US-0149175.
PR 18-JUN-1999;	99US-0139456.	PR 18-AUG-1999;	99US-0149426.
PR 18-JUN-1999;	99US-0139457.	PR 20-AUG-1999;	99US-0149722.
PR 18-JUN-1999;	99US-0139458.	PR 20-AUG-1999;	99US-0149723.
PR 18-JUN-1999;	99US-0139459.	PR 20-AUG-1999;	99US-0149929.
PR 18-JUN-1999;	99US-0139460.	PR 23-AUG-1999;	99US-0149902.
PR 18-JUN-1999;	99US-0139461.	PR 23-AUG-1999;	99US-0149930.
PR 18-JUN-1999;	99US-0139462.	PR 25-AUG-1999;	99US-0150566.
PR 18-JUN-1999;	99US-0139463.	PR 26-AUG-1999;	99US-0150884.
PR 18-JUN-1999;	99US-0139750.	PR 27-AUG-1999;	99US-0151065.
PR 18-JUN-1999;	99US-0139763.	PR 27-AUG-1999;	99US-0151066.
PR 21-JUN-1999;	99US-0139817.	PR 27-AUG-1999;	99US-0151080.
PR 22-JUN-1999;	99US-0139899.	PR 30-AUG-1999;	99US-0151303.
PR 23-JUN-1999;	99US-0140353.	PR 31-AUG-1999;	99US-0151438.
PR 23-JUN-1999;	99US-0140354.	PR 01-SEP-1999;	99US-0151930.
PR 24-JUN-1999;	99US-0140695.	PR 07-SEP-1999;	99US-0152363.
PR 28-JUN-1999;	99US-0140823.	PR 10-SEP-1999;	99US-0153070.
PR 29-JUN-1999;	99US-0140991.	PR 13-SEP-1999;	99US-0153758.
PR 30-JUN-1999;	99US-0141287.	PR 15-SEP-1999;	99US-0154018.
PR 01-JUL-1999;	99US-0141842.	PR 16-SEP-1999;	99US-0154039.
PR 01-JUL-1999;	99US-0142154.	PR 20-SEP-1999;	99US-0154779.
PR 02-JUL-1999;	99US-0142055.	PR 22-SEP-1999;	99US-0155139.
PR 06-JUL-1999;	99US-0142390.	PR 23-SEP-1999;	99US-0155486.
PR 08-JUL-1999;	99US-0142803.	PR 24-SEP-1999;	99US-0155659.
PR 09-JUL-1999;	99US-0142920.	PR 28-SEP-1999;	99US-0156458.
PR 12-JUL-1999;	99US-0142977.	PR 29-SEP-1999;	99US-0156596.
PR 13-JUL-1999;	99US-0143542.	PR 04-OCT-1999;	99US-0157117.
PR 14-JUL-1999;	99US-0143624.	PR 05-OCT-1999;	99US-0157753.
PR 15-JUL-1999;	99US-0144005.	PR 06-OCT-1999;	99US-0157865.
PR 16-JUL-1999;	99US-0144085.	PR 07-OCT-1999;	99US-0158029.
PR 16-JUL-1999;	99US-0144086.	PR 08-OCT-1999;	99US-0158237.
PR 19-JUL-1999;	99US-0144325.	PR 12-OCT-1999;	99US-0158368.
PR 19-JUL-1999;	99US-0144331.	PR 13-OCT-1999;	99US-0159293.
PR 19-JUL-1999;	99US-0144332.	PR 13-OCT-1999;	99US-0159294.
PR 19-JUL-1999;	99US-0144333.	PR 13-OCT-1999;	99US-0159295.
PR 19-JUL-1999;	99US-0144334.	PR 14-OCT-1999;	99US-0159329.
PR 19-JUL-1999;	99US-0144335.	PR 14-OCT-1999;	99US-0159330.
PR 20-JUL-1999;	99US-0144352.	PR 14-OCT-1999;	99US-0159331.
PR 20-JUL-1999;	99US-0144632.	PR 14-OCT-1999;	99US-0159637.
PR 21-JUL-1999;	99US-0144814.	PR 18-OCT-1999;	99US-0159638.
PR 21-JUL-1999;	99US-0145086.	PR 21-OCT-1999;	99US-0159584.
PR 21-JUL-1999;	99US-0145088.	PR 21-OCT-1999;	99US-0160741.
PR 22-JUL-1999;	99US-0145085.	PR 21-OCT-1999;	99US-0160767.
PR 22-JUL-1999;	99US-0145087.	PR 21-OCT-1999;	99US-0160768.
PR 22-JUL-1999;	99US-0145089.	PR 21-OCT-1999;	99US-0160770.
PR 23-JUL-1999;	99US-0145192.	PR 21-OCT-1999;	99US-0160814.
PR 23-JUL-1999;	99US-0145145.	PR 22-OCT-1999;	99US-0160815.
PR 23-JUL-1999;	99US-0145218.	PR 22-OCT-1999;	99US-0160980.
PR 23-JUL-1999;	99US-0145224.	PR 22-OCT-1999;	99US-0160981.
PR 26-JUL-1999;	99US-0145376.	PR 25-OCT-1999;	99US-0161404.

PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 37.5%; Score 3; DB 21; Length 121;
Best Local Similarity 100.0%; Pred. No. 6.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ayxg 8
|||
Db 57 ayxg 60

RESULT 6
AAW20262
ID AAW20262 standard; Protein: 185 AA.
AC AAW20262;
XX 30-JUL-1997 (first entry)
DT H. pylori secreted or periplasmic protein 23594833.aa.
DE
XX Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT Misc-difference 2
FT /label= "unknown
FT /note= "encoded by YAT"
FT Misc-difference 101
FT /label= "unknown
FT /note= "encoded by RAA"
FT Misc-difference 142
FT /label= "unknown
FT /note= "encoded by RAA"
FT Misc-difference 149
FT /label= "unknown
FT /note= "encoded by CSG"
FT Misc-difference 157
FT /label= "unknown
FT /note= "encoded by GAK"
FT Misc-difference 158
FT /label= "unknown
FT /note= "encoded by WAA"
XX
PN W09640893-A1.
XX 19-DEC-1996.
PD 06-JUN-1996; 96WO-US09122.
XX 01-APR-1996; 96US-0630405.
PR 07-JUN-1995; 95US-0487032.
XX
XX (ASTR) ASTRA AB.
XX
XX Berghindh OT, Smith D, Meligaerd BL;
XX
DR WPI, 1997-052306/05.
DR N-PSDB; AAT67760.
XX
XX Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori

PT infection, and to detect Helicobacter
XX
XX Claim 72; Page 466; 1481pp: English.
PS
XX This sequence is a H. pylori secreted or periplasmic protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
XX
SQ Sequence 185 AA;

Query Match 37.5%; Score 3; DB 18; Length 185;
Best Local Similarity 100.0%; Pred. No. 9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ayxg 8
|||
Db 140 ayxg 143

RESULT 7
AAW24622
ID AAW24622 standard; Protein: 185 AA.
AC AAW24622;
XX 11-AUG-1997 (first entry)
DT H. pylori secreted or periplasmic protein 23594833.aa.
DE
XX Transmembrane; cytoplasmic; cell envelope; flagella; transport;
KW secreted; periplasmic; chronic gastritis; duodenal ulcer disease;
KW activator; inhibitor; bacterial life cycle; vaccine; immunise;
KW detection; antisense; inhibition.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT Misc-difference 2
FT /label= "unknown
FT /note= "encoded by YAT"
FT Misc-difference 101
FT /label= "unknown
FT /note= "encoded by RAA"
FT Misc-difference 142
FT /label= "unknown
FT /note= "encoded by RAA"
FT Misc-difference 149
FT /label= "unknown
FT /note= "encoded by CSG"
FT Misc-difference 157
FT /label= "unknown
FT /note= "encoded by GAK"
FT Misc-difference 158
FT /label= "unknown
FT /note= "encoded by WAA"
XX
PN W09719098-A1.
XX 29-MAY-1997.
PD 15-NOV-1996; 96WO-US18542.
XX 17-NOV-1995; 95US-0561469.
XX
XX (ASTR) ASTRA AB.
XX
XX Smith DH;

XX WPI: 1997-298052/27.
DR N-PSDB: AAG77440.
XX Helicobacter pylori nucleic acid sequences and related proteins -
PT used for diagnostics and therapeutics
PS Claim 10; Page 153; 1481pp; English.
XX This sequence is a H. pylori secreted or periplasmic protein.
CC Helicobacter pylori has been strongly linked to chronic gastritis and
CC duodenal ulcer disease. The nucleic acid sequences of the invention
CC are used to evaluate compounds, especially activators or inhibitors of
CC bacterial life cycle, for the ability to bind an H. pylori nucleic acid
CC sequence. The nucleic acid sequences, and corresponding proteins, are
CC also useful for generating vaccines for immunising subjects against H.
CC pylori or for use in detecting the presence of Helicobacter species in
CC a sample. Antisense nucleic acid sequences of these sequences are
CC used to inhibit expression of a gene from Helicobacter species. H.
CC pylori whole genomic DNA was isolated and nebulised to a median size of
CC 2000 bp. Purified DNA fragments were blunt-ended and ligated to unique
CC BstXI-linker adapters in 100-1000 fold molar excess. These linkers are
CC complementary to the BstXI-cut PMPX vectors, while the overhang is not
CC self-complementary. Therefore the linkers will not concatemerise nor
CC will the cut vector re-ligate itself easily. The linker-adaptor inserts
CC were ligated to each of the 20 PMPX vectors to construct a series of
CC shotgun subclone libraries. The purified DNA samples were then
CC sequenced.
CC Note: The ORF/protein reference number for this sequence was obtained
CC from the related specification, WO9640893.
SQ Sequence 185 AA:
OY 5 ayxg 8 37.5%; Score 3; DB 18; Length 185;
||| Best Local Similarity 100.0%; Pred. No. 9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 140 ayxg 143
RESULT 8
AAR10204
ID AAR10204 standard; Protein: 275 AA.
XX AAR10204:
AC
XX 27-MAR-1991 (first entry)
DT
XX
XX Mutant subtilisin polypeptide (1).
DE
XX
XX Mutant; subtilisin; detergent; serine protease.
KM
XX
XX Key Location/Qualifiers
FH Misc-difference 218..218
FT /label: S, D
FT
XX
XX US4980288-A.
PN
XX
XX 25-DEC-1990.
PD
XX
XX 14-DEC-1987; 87US-0143949.
PF
XX
XX 14-DEC-1987; 87US-0143949.
PR
XX 12-FEB-1986; 86US-0828345.
PR
XX (GENE-) GENEX CORP.
PA
XX Bryan PN, ROLLENCE MJ, Pantollano MW;
PI
XX WPI: 1991-021675/03.
DR

XX Mutant subtilisin poly-peptide(s) with increased thermal
PT stability - obtd. by muta-genesis of subtilisin gene used in 11q.
PT detergent compns.
XX
PS Claim 1; Page 16; 18pp; English.
XX NB: to construct sequences AAR10204-13, the B. subtilis wild-type
CC sequence was retrieved from the GENESQ database (AAP90095; J01137972-A)
CC and amino acid residues altered according to the mutations described
CC in the specification (Ser/Asp-Asn218)
CC The mutant subtilisin does not lose activity as rapidly when stored
CC in soln. with detergents or when subjected to high heat during use
CC in cleaning. It is used for the removal of proteinaceous stains on
CC fabric. Other stabilising mutants may be introduced into the
CC protein. The stabilising mutations may also be applied to other
CC homologous serine proteases.
XX
SQ Sequence 275 AA:
OY 5 ayxg 8 37.5%; Score 3; DB 12; Length 275;
||| Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 216 ayxg 219
RESULT 9
AAG75013
ID AAG75013 standard; Protein: 277 AA.
XX
XX AAG75013:
AC
XX 03-SEP-2001 (first entry)
DT
XX
XX Human colon cancer antigen protein SEQ ID NO:5777.
DE
XX
XX Human: colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 14.
KW
XX
XX Homo sapiens.
OS
XX
XX WO200122920-A2.
PN
XX
XX 05-APR-2001.
PD
XX
XX 28-SEP-2000; 2000WO-US26524.
PF
XX
XX 29-SEP-1999; 99US-0157137.
PR
XX 03-NOV-1999; 99US-0163280.
PR
XX
XX (HUMA-) HUMAN GENOME SCT INC.
PA
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI
XX
XX WPI: 2001-235357/24.
DR
XX
XX N-PSDB: AAH34418.
DR
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
PT
XX
PS Claim 11; Page 7294-7295; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders

CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of p by expressing
CC inactive proteins or to supplement the patients own production of p.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAF1506 to AAF1514 to
CC and AAF7789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 277 AA;

Query Match 37.5%; Score 3; DB 22; Length 277;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ayyg 8
|||
Db 89 ayyg 92

RESULT 10
AAB56744 standard; Protein; 284 AA.
XX
AC AAB56744;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human prostate cancer antigen protein sequence SEQ ID NO:1322.
XX
XX Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
XX neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
XX vulnery; gastrointestinal; nephrotoxic; antiinfective; gynaecological;
XX antibacterial; gene therapy; neural; immune; reproductive; renal;
XX gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
XX wound; infectious disease.
XX
XX Homo sapiens.
XX OS
XX MO20005174-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000MO-US05988.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HOMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
XX Rosen CA, Ruben SM;
XX MPI: 2000-587513/55.
XX N-PSDB; AAF15947.
XX
XX prostate cancer associated gene sequences, referred to as prostate
XX cancer antigens, useful for treatment, prevention, and diagnosis of
XX disorders such as prostate cancer -
XX
XX Claim 11; Page 1743-1744; 2338pp; English.
XX
XX AAF15566 to AAF16505 encode the human prostate cancer associated
XX proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
XX The prostate cancer antigens can have neuroprotective, cytostatic,
XX cardioactive, immunomodulatory, muscular, vulnerary, gastrointestinal,
XX nephrotoxic, antiinfective, gynaecological and antibacterial activities,
XX and can be used in gene therapy. The prostate cancer antigen
XX polynucleotides may be used for detection of prostate cancer, chromosome

CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF1506 to AAF1514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 284 AA;

Query Match 37.5%; Score 3; DB 21; Length 284;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ayyg 8
|||
Db 263 ayyg 266

RESULT 11
AAR12129 standard; Protein; 401 AA.
XX
AC AAR12129;
XX
DT 01-AUG-1991 (first entry)
XX
DE ORF 1 of IgG light chain variable region clone.
XX
XX Immunoglobulin G; light chain; variable region; duplication;
XX passive immunity; group B streptococci.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FH 224..336
XX FT Peptide /label="L.V region
XX FT /note="last 3 amino acids of leader and variable
XX FT region"
XX
XX WO9106305-A.
XX
XX 16-MAY-1991.
XX
XX 06-NOV-1990; 90MO-US06426.
XX
XX 07-NOV-1989; 89US-0432700.
XX
XX (BRIM) BRISTOL-MYERS SQUIB.
XX
XX Shuford WW, Harris LJ, Rafi HV;
XX MPI: 1991-163947/22.
XX N-PSDB; AAQ11879.
XX
XX Oligomeric immunoglobulin(s) with high avidity for antigen(s) -
XX PT formed by duplicating esp. variable region of light chain of IgG
XX PT class
XX
XX Example 4; Fig 17; 104pp; English.
XX
XX This sequence is derived from the nucleotide sequence encoding the
XX light chain variable region. The "x" residues represent nonsense
XX codons. The coding sequence has been translated in all 3 reading
XX frames (see also AAR12130 and AAR12131). The L.V region is duplicated
XX in so-called "aberrant" light chains (see AAQ11878), conferring
XX increased avidity on antibodies comprising such aberrant chains.
XX See also AAQ11880.
XX
XX Sequence 401 AA;

Query Match 37.5%; Score 3; DB 12; Length 401;
 Best Local Similarity 100.0%; Pred. No. 1.3e+03;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 5 ayxg 8
 ||||
 Db 27 ayxg 30

RESULT 12

AAW84298
 ID AAW84298 standard; Protein: 489 AA.

AC AAW84298;

DT 25-MAR-1999 (first entry)

DE Consensus sequence of GDNFR and GRR2 and GRR3 proteins.

XX Glial cell-line derived neurotrophic factor receptor;

KW GDNFR; glial cell line-derived neurotrophic factor; GDNF;

KW neurturin; signal transduction; dopaminergic nerve cell;

KW Parkinson's disease; Alzheimer's disease; amyotrophic lateral sclerosis;

KW neurological disorder; diabetes; glaucoma; sensory neuron;

KW retinal ganglion cell degeneration; sensory neuropathy;

KW retinopathy; gene therapy; GDNFR-related protein; GRR2; GRR3.

XX Synthetic.

OS Key Location/Qualifiers

FT Misc-difference 1..489 /note= "x-not specified"

PN W09854213-A2.

PD 03-DEC-1998.

PE 27-APR-1998; 98WO-US08486.

PR 30-MAY-1997; 97US-0866354.

XX (AMGE-) AMGEN INC.

PI Fox GM, Jing S, Wen D;

WT: 1999-080806/07.

XX New isolated glial cell line-derived neurotrophic factor receptors -

PT used to develop products for treating e.g. improperly functioning

PT dopaminergic nerve cells, Parkinson's disease, Alzheimer's disease

PT or amyotrophic lateral sclerosis

XX Claim 51; Fig 26; 318bp; English.

CC The present sequence represents a consensus sequence of glial

CC cell-line derived neurotrophic factor receptor (GDNFR) protein and

CC GDNFR-related (GRR) proteins GRR2 and GRR3. The proteins have similar

CC functions. GDNFR proteins are functionally characterised by the ability

CC to bind glial cell line-derived neurotrophic factor (GDNF) and/or

CC neurturin specifically, and to act as part of a molecular complex which

CC mediates or enhances the signal transduction affects of GDNF and/or

CC neurturin. The proteins can be used for treating improperly functioning

Query Match 37.5%; Score 3; DB 20; Length 489;
 Best Local Similarity 100.0%; Pred. No. 1.8e+03;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 5 ayxg 8
 ||||
 Db 305 ayxg 308

RESULT 13

AAW84183
 ID AAW84183 standard; Protein: 498 AA.

AC AAW84183;

DT 25-MAR-1999 (first entry)

DE Consensus sequence of rat and human GDNFR and GRR proteins.

XX Glial cell-line derived neurotrophic factor receptor;

KW GDNFR; glial cell line-derived neurotrophic factor; GDNF;

KW neurturin; signal transduction; dopaminergic nerve cell;

KW Parkinson's disease; Alzheimer's disease; amyotrophic lateral sclerosis;

KW neurological disorder; diabetes; glaucoma; sensory neuron;

KW retinal ganglion cell degeneration; sensory neuropathy;

KW retinopathy; gene therapy; GDNFR-related protein; GRR.

XX Synthetic.

OS Rattus sp.

OS Homo sapiens.

FT Key Location/Qualifiers

FT Misc-difference 1..498 /note= "x-not specified"

PN W09854213-A2.

PD 03-DEC-1998.

PE 27-APR-1998; 98WO-US08486.

PR 30-MAY-1997; 97US-0866354.

XX (AMGE-) AMGEN INC.

PI Fox GM, Jing S, Wen D;

WT: 1999-080806/07.

XX New isolated glial cell line-derived neurotrophic factor receptors -

PT used to develop products for treating e.g. improperly functioning

PT dopaminergic nerve cells, Parkinson's disease, Alzheimer's disease

PT or amyotrophic lateral sclerosis

XX Claim 51; Fig 19; 318bp; English.

CC The present sequence represents a consensus sequence of rat and

CC human glial cell-line derived neurotrophic factor receptor (GDNFR)

CC protein and GDNFR-related (GRR) proteins. The proteins have similar

CC functions. GDNFR proteins are functionally characterised by the ability

CC to bind glial cell line-derived neurotrophic factor (GDNF) and/or

CC neurturin specifically, and to act as part of a molecular complex which

mediates or enhances the signal transduction affects of GDNF and/or

neurturin. The proteins can be used for treating improperly functioning

dopaminergic nerve cells, Parkinson's disease, Alzheimer's disease or

amyotrophic lateral sclerosis. They can also be used for treating

neurological disorders associated with diabetes, glaucoma or other

diseases and conditions involving retinal ganglion cell degeneration,

#sensory neuropathy caused by injury to, insults to, or degeneration of,

sensory neurons, pathological conditions, or disease or injury-related

retinopathies. The products can also be used for detection, diagnosis,

drug screening and gene therapy.

```
XX SQ Sequence 498 AA:
Query Match 37.5%; Score 3; DB 20; Length 498;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 ayxg 8
   ||||
Db 312 ayxg 315

RESULT 14
AAM40758
ID AAM40758 standard; Protein; 500 AA.
XX AAM40758;
XX
XX 22-OCT-2001 (first entry)
XX
XX Human polypeptide SEQ ID NO 5689.
XX
XX Human; noctropic; immunosuppressant; cytosolic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokineic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia.
XX
XX Homo sapiens.
XX
XX WO200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
XX N-PSDB; AA159914.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Example 2; SEQ ID NO 5689; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AAM42213) with noctropic,
XX immunosuppressant and cyostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
```

```
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX SQ Sequence 500 AA:
Query Match 37.5%; Score 3; DB 22; Length 500;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 ayxg 8
   ||||
Db 332 ayxg 335

RESULT 15
AA38787
ID AA38787 standard; Protein; 513 AA.
XX AA38787;
XX
XX 08-OCT-1999 (first entry)
XX
XX Neisseria meningitidis strain A antigen encoded by ORF139.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
XX Neisseria meningitidis.
XX
XX WO9924578-A2.
XX
XX 20-MAY-1999.
XX
XX 09-OCT-1998; 98WO-IB01665.
XX
XX 01-SEP-1998; 98GB-0019016.
XX 06-NOV-1997; 97GB-0023516.
XX 14-NOV-1997; 97GB-0024190.
XX 18-NOV-1997; 97GB-0024386.
XX 27-NOV-1997; 97GB-0025158.
XX 10-DEC-1997; 97GB-0026147.
XX 14-JAN-1998; 98GB-0000759.
XX
XX (CHIR-) CHIRON SPA.
XX
XX Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX
XX WPI: 1999-327407/27.
XX N-PSDB; AA122222.
XX
XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
XX diagnosis, treatment and prevention of infection
XX
XX Claim 4; Page 331; 524pp; English.
XX
XX Amino acid sequences AA38499-Y38944 represent Neisseria meningitidis
XX and N. gonorrhoeae antigenic proteins. They are encoded by open
XX reading frames (ORFs) AA21972-212358. The antigenic proteins,
XX their fragments, their nucleic acids and antibodies are used for
XX diagnosis, prevention (as vaccines) or treatment of Neisseria
XX infections, such as meningitis, septicaemia and gonorrhea. Both
XX organisms are closely related. Fragments of the nucleic acids
XX are useful as hybridisation probes and antisense reagents.
XX
XX SQ Sequence 513 AA:
Query Match 37.5%; Score 3; DB 20; Length 513;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

OY 5 ayxg 8
1111
Db 470 ayxg 473

RESULT 16
AAP80575
ID AAP80575 standard; protein; 515 AA.
XX AAP80575;
XX 08-NOV-1990 (first entry)
XX Mutated alpha-amylase of Bacillus stearothersophilus.
XX B.stearothersophilus alpha-amylase; base substitutions.
XX Bacillus stearothersophilus.
OS
XX
XX Key Location/Qualifiers
FH Misc-difference 13..13 /label=V, E
FT /note="Wild-Type is E"
FT Misc-difference 15..15 /label=S, F, Y
FT /note="Wild-Type is Y"
FT Misc-difference 18..18 /label=A, V, D
FT /note="Wild-Type is D"
FT Misc-difference 19..19 /label=A, V, D
FT /note="Wild-Type is D"
FT Misc-difference 21..21 /label=P, S, T
FT /note="Wild-Type is T"
FT Misc-difference 22..22 /label=F, L
FT /note="Wild-Type is L"
FT Misc-difference 24..24 /label=P, T
FT /note="Wild-Type is T"
FT Misc-difference 25..25 /label=N, K
FT /note="Wild-Type is K"
FT Misc-difference 31..31 /label=Y, N
FT /note="Wild-Type is N"
FT Misc-difference 32..32 /label=I, N
FT /note="Wild-Type is N"
FT Misc-difference 33..33 /label=F, L
FT /note="Wild-Type is L"
FT Misc-difference 34..34 /label=C, S
FT /note="Wild-Type is S"
FT Misc-difference 48..48 /label=T, K
FT /note="Wild-Type is K"
FT Misc-difference 51..51 /label=R, S
FT /note="Wild-Type is S"
FT Misc-difference 53..53 /label=R, S
FT /note="Wild-Type is S"
FT Misc-difference 54..54 /label=A, V, D
FT /note="Wild-Type is D"
FT Misc-difference 57..57 /label=S, C, Y
FT /note="Wild-Type is Y"
FT Misc-difference 60..60

FT /label=F, Y
FT /note="Wild-Type is Y"
FT Misc-difference 61..61 /label=V, D
FT /note="Wild-Type is D"
FT Misc-difference 63..63 /label=C, F, Y
FT /note="Wild-Type is Y"
FT Misc-difference 67..67 /label=D, E
FT /note="Wild-Type is E"
FT Misc-difference 126..126 /label=H, R
FT /note="Wild-Type is R"
FT Misc-difference 127..127 /label=T, N
FT /note="Wild-Type is N"
FT Misc-difference 129..129 /label=G, E
FT /note="Wild-Type is E"
FT Misc-difference 134..134 /label=S, Y
FT /note="WT is Y"
FT Misc-difference 137..137 /label=P, Q
FT /note="WT is Q"
FT Misc-difference 138..138 /label=T, A
FT /note="WT is A"
FT Misc-difference 250..250 /label=F, Y
FT /note="WT is Y"
FT Misc-difference 254..254 /label=P, Q
FT /note="WT is Q"
FT Misc-difference 255..255 /label=P, T
FT /note="WT is T"
FT Misc-difference 257..257 /label=M, T, Q, K
FT /note="WT is K"
FT Misc-difference 261..261 /label=S, A, T
FT /note="WT is T"
FT Misc-difference 264..264 /label=D, V, E
FT /note="WT is E"
FT Misc-difference 265..265 /label=F, S, Y
FT /note="WT is Y"
FT Misc-difference 277..277 /label=F, I
FT /note="WT is I"
FT Misc-difference 291..291 /label=F, L
FT /note="WT is L"
FT Misc-difference 331..331 /label=A, D
FT /note="WT is D"
FT Misc-difference 332..332 /label=P, T
FT /note="WT is T"
FT Misc-difference 333..333 /label=A, E
FT /note="WT is E"
FT Misc-difference 335..335 /label=C, G
FT /note="WT is G"
FT Misc-difference 336..336 /label=P, H, L, Q
FT /note="WT is Q"
FT Misc-difference 337..337 /label=T, A

FT /note="WT is A"
FT Misc-difference 343..343
FT /label=A, D
FT /note="WT is D"
FT Misc-difference 347..347
FT /label=N, K
FT /note="WT is K"
FT Misc-difference 351..351
FT /label=S, Y
FT /note="WT is Y"
FT Misc-difference 354..354
FT /label=L, I
FT /note="WT is I"
FT Misc-difference 369..369
FT /label=S, Y
FT /note="WT is Y"
FT Misc-difference 370..370
FT /label=S, Y
FT /note="WT is Y"
FT Misc-difference 414..414
FT /label=V, E
FT /note="WT is E"
FT Misc-difference 419..419
FT /label=N, K
FT /note="WT is K"
FT Misc-difference 428..428
FT /label=F, I
FT /note="WT is I"
FT Misc-difference 435..435
FT /label=F, S
FT /note="WT is S"
FT Misc-difference 439..439
FT /label=F, Y
FT /note="WT is Y"
FT Misc-difference 442..442
FT /label=N, Q, K
FT /note="WT is K"
FT Misc-difference 443..443
FT /label=H, Q
FT /note="WT is Q"
FT Misc-difference 489..489
FT /label=P, T
FT /note="WT is T"
XX
XX EP285123-A.
XX
XX PD 05-MAY-1988.
XX
XX PE 30-MAR-1988; 88EP-0105163.
XX
XX PR 03-APR-1987; 87US-0034819.
XX
XX PA (SUSO) SUOMEN SOKERI OY.
XX
XX PI Lehtovaara P, Knowles J, Kolvula A, Bamford J, Reinikainen T;
XX
XX DR WPI; 1988-279927/40.
XX
XX PT Introducing random point mutations into nucleic acids -
XX PT by prepn of single stranded template, annealing a primer, elongation,
XX PT misincorporation, completion of molecules and screening.
XX

Query Match 37.5%; Score 3; DB 9; Length 515;
Best Local Similarity 100.0%; Pred.No.1.8e+03;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 axyg 8
|||
Db . 46 axyg 49

RESULT 17 .

AAW17569
XX AAW17569 standard; protein: 688 AA.
AC AAW17569;
XX
XX 30-JUN-1997 (first entry)
XX
XX DE Bacillus CGase variant beta-cyclodextrin #1.
XX
XX KM Cyclomaltodextrin glucanotransferase; CGase; enzyme; Bacillus circulans;
KM thermocaneobacter; starch; cyclomaltodextrin; cyclodextrin; pesticide;
KM transglycosylation; cyclic glucose oligomer; glucose oligosaccharide;
KM cholesterol removal; food stabiliser; vitamin stabiliser; fungicide;
KM plastic laminate; biodegradable plastic; mutein.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FH Modified-site 21
FT /label= F21Y
FT
FT Modified-site 47
FT /label= R47X
FT /note= "X= Ala, Gln, His, Arg, Leu"
FT Modified-site 87
FT /label= I87H
FT
FT Modified-site 88
FT /label= N88X
FT /note= "X= Pro, Asn, Lys, His"
FT Modified-site 89
FT /label= Y89X
FT /note= "X= Asp, Gly, Ala, Lys, Arg, Pro, Tyr, Glu or
FT absent"
FT
FT Modified-site 90
FT /label= S90X
FT /note= "X= Gly, Ala, Ser"
FT Modified-site 91
FT /label= G91X
FT /note= "X= Ala, Val, Ser, Gly, Thr"
FT Misc-difference 92
FT /note= "possible insertion of Ala, Val, Gly, Phe, or Tyr"
FT 93
FT /label= V92G
FT
FT Modified-site 94
FT /label= N93X
FT /note= "X= Gly, Asn, His, Thr, or absent"
FT Modified-site 95
FT /label= N94X
FT /note= "X= Gln, Lys, Arg, Trp, Phe, Asn, Ser, or absent"
FT Modified-site 99
FT /label= H98X
FT /note= "X= Gly, Ala"
FT Modified-site 102
FT /label= W101X
FT /note= "X= Gly, Ala"
FT Modified-site 136
FT /label= D135L
FT
FT Modified-site 141
FT /label= H140X
FT /note= "X= Ala, Arg, Asn"
FT Modified-site 146
FT /label= S145X
FT /note= "X= Ala, Glu, Trp, Leu"
FT Misc-difference 147
FT /note= "insertion of Pro, Ala, Ile, Gln, or Ser"
FT 148
FT /label= S146X
FT /note= "X= Pro, Ala, Phe, Gln, Ser, Trp, Ile, Arg, Glu,
FT Lys, Asp, Asn, or absent"
FT Modified-site 149
FT /label= D147X
FT /note= "X= Ala, Leu, Ile, Phe, Trp, Gly, Tyr, Arg, Asp,
FT Thr, or absent"
FT Modified-site 150

PT	Modified-site	/label_ Q148X /note_ "X" Gly, Asn, Ala, Gln"
PT	Modified-site	/label_ P149W /label_ S150A
PT	Modified-site	/label_ Y167X /note_ "X" Ala, Phe, Tyr"
PT	Modified-site	/label_ T168S /label_ N178Y
PT	Modified-site	/label_ G179X /note_ "X" Gly, Ser, Asn, Asp"
PT	Modified-site	/label_ G180X /note_ "X" Gly, Ser, Asn, Asp"
PT	Modified-site	/label_ F183X /note_ "X" Phe, Trp, Tyr, Ala"
PT	Modified-site	/label_ T185X /note_ "X" Pro, His, Arg, Glu, Asp"
PT	Modified-site	/label_ K192R /label_ N193X
PT	Modified-site	/note_ "X" Gly, Ala, Asn"
PT	Modified-site	/label_ L195X /note_ "X" Leu, Ile, Trp, Tyr, Phe"
PT	Modified-site	/note_ "X" Ala, Asp, Asn, Ser"
PT	Modified-site	/label_ L197X /note_ "X" Asp, Glu, Leu"
PT	Modified-site	/label_ K232X /note_ "X" Lys, Gln, Leu"
PT	Modified-site	/label_ H233X /note_ "X" His, Gln, Asn, Ile"
PT	Modified-site	/label_ F259X /note_ "X" Phe, Trp, Tyr, Ala"
PT	Modified-site	/label_ E264Q /label_ N326X
PT	Modified-site	/note_ "X" Gln, Phe, Leu"
PT	Modified-site	/label_ T370N /label_ D371X
PT	Modified-site	/note_ "X" Ala, Asp, Ser, Asn, Gly"
PT	Modified-site	/label_ D373X /note_ "X" Asp, Glu, Tyr"
PT	Modified-site	/label_ R375X /note_ "X" Ala, Pro, Gly, Arg, Lys"
PT	Modified-site	/label_ L600X /note_ "X" unspecified amino acid"
PD	24-OCT-1996.	
XX	W09633267-A1.	
XX	22-APR-1996;	96WO-DK00179.

XX	PR	16-NOV-1995;	95DK-0001281.	
XX	PR	21-APR-1995;	95DK-0000477.	
XX	PR	17-OCT-1995;	95DK-0001173.	
XX	PA	(NOVO)	NOVO-NORDISK AS.	
XX	PI	Andersen C,	Dijkhuizen L, Dijkstra BW, Von Der Osten C;	
XX	DR	WPI: 1996-485774/48.		
XX	PT	New variants of cyclomalto:	dextrin glucanotransferase (CGTase) -	
XX	PT	have altered substrate binding,	useful for prodn. of cyclodextrins)	
XX	PT	or linear oligosaccharide(s),	opt. formed in situ in e.g. baked	
XX	PS	goods		
XX	PS	Claim 31; Page -;	161pp; English.	
XX	XX	AAW17568-W17579, and AAW17592-W17605	represent mutant versions of the	
XX	CC	cyclomalto:dextrin glucanotransferase (CGTase)	of Bacillus circulans	
XX	CC	strain 251 (see AAW06773 for wild type sequence).	CGTase catalyses the	
XX	CC	conversion of starch and similar substrates	into cyclomalto:dextrins	
XX	CC	(also known as cyclodextrins) via an intramolecular	transglycosylation	
XX	CC	reaction. Cyclodextrins (CD) are cyclic glucose	oligomers with a	
XX	CC	hydrophobic internal cavity that form inclusion	complexes with many	
XX	CC	small hydrophobic molecules. These CGTase	mutants have a modified	
XX	CC	substrate binding and/or product selectivity,	compared to this sequence.	
XX	CC	The mutants are created using primer mutagenesis	to modify the gene	
XX	CC	encoding this sequence. These sequences have	greater product selectivity	
XX	CC	and/or reduced product inhibition (better	yields) than wild-type CGTase.	
XX	CC	These mutant sequences are used to manufacture	the 6 main CD types	
XX	CC	(comprising 6-11 glucose units), or linear	2-12 glucose oligosaccharides,	
XX	CC	optionally in situ, e.g. during production	of baked goods, to stabilise	
XX	CC	chemicals during their manufacture and in	detergents. CD are known for	
XX	CC	their usefulness in foods, e.g. as a bread-	improving agent, to	
XX	CC	encapsulate/stabilise/solubilise vitamins,	dyes, pharmaceuticals,	
XX	CC	pesticides or fungicides, to bind/remove	lipophilic compounds such as	
XX	CC	cholesterol (e.g. in egg yolk or butter),	in plastic laminates, films	
XX	CC	etc. and to make biodegradable plastics.		
XX	XX	Sequence 688 AA;		
SQ				
		Query Match	37.5%; Score 3; DB 17; Length 688;	
		Best Local Similarity	100.0%; Pred. No. 2.2e+03;	
		Matches 4; Conservative 0; Mismatches 0;	Indels 0; Gaps 0;	
OY		5 ayxg 8		
Db		97 ayxg 100		
		RESULT 18		
	ID	AAW17577		
	ID	AAW17577 standard; protein; 688 AA.		
	AC	AAW17577;		
XX	XX	30-JUN-1997 (first entry)		
XX	DE	Bacillus CGTase variant beta-cyclodextrin #3.		
XX	XX	Cyclomalto:dextrin glucanotransferase; CGTase; enzyme; Bacillus circulans;		
XX	KW	thermoanaerobacter; starch; cyclomalto:dextrin; cyclodextrin; pesticide;		
XX	KW	transglycosylation; cyclic glucose oligomer; glucose oligosaccharide;		
XX	KW	cholesterol removal; food stabiliser; vitamin stabiliser; fungicide;		
XX	KW	plastic laminate; biodegradable plastic; muten.		
XX	OS	Synthetic.		
XX	XX			
XX	PH	Key	Location/Qualifiers	
XX	FT	Modified-site	21	
XX	FT	/label= F21Y		

FT	Modified-site	47	/label= R47X
FT			/note= "X= Ala, Gln, His, Arg, Leu"
FT	Modified-site	87	/label= I87H
FT	Modified-site	88	/label= N88X
FT			/note= "X= Asn, Lys, His"
FT	Modified-site	89	/label= Y89X
FT			/note= "X= Gly, Ala, Lys, Arg, Pro, Tyr, Glu or absent"
FT	Modified-site	90	/label= S90X
FT			/note= "X= Gly, Ala"
FT	Modified-site	91	/label= G91X
FT			/note= "X= Ala, Val, Ser, Gly"
FT	Misc-difference	92	/note= "possible insertion of Ala, Val, Gly, or Tyr"
FT			/note= "X= Gly, Ala"
FT	Modified-site	94	/label= N93X
FT			/note= "X= Asn, His, Thr, or absent"
FT	Modified-site	95	/label= N94X
FT			/note= "X= Gln, Lys, Arg, Trp, Phe, Asn, or absent"
FT	Modified-site	99	/label= H98X
FT			/note= "X= Gly, Ala"
FT	Modified-site	102	/label= W101X
FT			/note= "X= Gly, Ala"
FT	Modified-site	136	/label= D135L
FT			/label= H140X
FT	Modified-site	141	/label= H140X
FT			/note= "X= Ala, Arg, Asn"
FT	Modified-site	146	/label= S145X
FT			/note= "X= Ala, Glu, Trp, Leu"
FT	Misc-difference	147	/note= "insertion of Pro, Ala, Ile, Gln, or Ser"
FT			/label= S146X
FT	Modified-site	148	/label= S146X
FT			/note= "X= Pro, Ala, Phe, Gln, Ser, Trp, Ile, Arg, Lys, Asp, Asn, or absent"
FT	Modified-site	149	/label= D147X
FT			/note= "X= Ala, Leu, Ile, Phe, Trp, Gly, Tyr, Arg, Asp, or absent"
FT	Modified-site	150	/label= Q148X
FT			/note= "X= Gly, Asn, Ala, Gln"
FT	Modified-site	151	/label= P149W
FT	Modified-site	152	/label= S150A
FT	Modified-site	169	/label= Y167X
FT			/note= "X= Ala, Phe"
FT	Modified-site	170	/label= T168S
FT			/label= G179X
FT	Modified-site	182	/label= G180X
FT			/note= "X= Ser, Asn, Asp"
FT	Modified-site	185	/label= F183X
FT			/note= "X= Trp, Tyr, Ala"
FT	Modified-site	187	/label= T185X
FT			/note= "X= Pro, His, Arg, Glu, Asp"

FT	Modified-site	195	/label= N193X
FT			/note= "X= Gly, Ala"
FT	Modified-site	197	/label= L195X
FT			/note= "X= Ieu, Ile, Trp, Tyr"
FT	Modified-site	198	/label= D196X
FT			/note= "X= Ala, Gly, Asn, Ser"
FT	Modified-site	199	/label= L197X
FT			/note= "X= Asp, Glu"
FT	Modified-site	234	/label= K232X
FT			/note= "X= Gln, Leu"
FT	Modified-site	235	/label= H233X
FT			/note= "X= Gln, Asn, Ile"
FT	Modified-site	261	/label= F259X
FT			/note= "X= Phe, Trp, Ala"
FT	Modified-site	266	/label= E264Q
FT			/label= N326X
FT	Modified-site	372	/label= T370N
FT			/label= D371X
FT	Modified-site	375	/label= R375X
FT			/note= "X= Ala, Pro, Gly, Lys"
FT	Modified-site	602	/label= L600X
FT			/note= "X= unspecified amino acid"
FT	Modified-site	W09633267-A1.	
FT			24-OCT-1996.
FT	Modified-site	22-APR-1996;	96NO-DK00179.
FT			16-NOV-1995;
FT			21-APR-1995;
FT			17-OCT-1995;
FT			(NOVO) NOVO-NORDISK AS.
FT			Andersen C, Dijkhuizen L, Dijkstra BW, Von Der Osten C;
FT			WPI; 1996-485774/48.
FT			New variants of cyclo,malto:dextrin glucanotransferase (CGTase) -
FT			have altered substrate binding, useful for prodn. of cyclodextrin(s)
FT			or linear oligosaccharide(s), opt. formed in situ in e.g. baked
FT			goods
FT			Claim 40; Page -; 161pp; English.
FT			AAW17568-W17579, and AAW17592-W17605 represent mutant versions of the
FT			cyclomalto-dextrin glucanotransferase (CGTase) of Bacillus circulans
FT			strain 251 (see AAW06773 for wild type sequence). CGTase catalyses the
FT			conversion of starch and similar substrates into cyclomalto-dextrins
FT			(also known as cyclodextrins) via an intramolecular transglycosylation
FT			reaction. Cyclodextrins (CD) are cyclic glucose oligomers with a
FT			small hydrophobic internal cavity that form inclusion complexes with many
FT			small hydrophobic molecules. These CGTase mutants have a modified
FT			substrate binding and/or product selectivity, compared to this sequence.
FT			The mutants are created using primer mutagenesis to modify the gene

CC encoding this sequence. These sequences have greater product selectivity
CC and/or reduced product inhibition (better yields) than wild-type CGTase.
CC These mutant sequences are used to manufacture the 6 main CD types
CC (comprising 6-11 glucose units), or linear 2-12 glucose oligosaccharides,
CC optionally in situ, e.g. during production of baked goods, to stabilise
CC chemicals during their manufacture and in detergents. CD are known for
CC their usefulness in foods, e.g. as a bread-improving agent, to
CC encapsulate/stabilise/solubilise vitamins, dyes, pharmaceuticals,
CC pesticides or fungicides, to bind/remove lipophilic compounds such as
CC cholesterol (e.g. in egg yolk or butter), in plastic laminates, films
CC etc. and to make biodegradable plastics.
XX
XX
SO Sequence 688 AA;

Query Match 37.5%; Score 3; DB 17; Length 688;
Best Local Similarity 100.0%; Pred. No. 2.2e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 aysg 8
1111
Db 97 aysg 100

RESULT 19
AAW17573
ID AAW17573 standard; protein; 688 AA.
XX
XX AAW17573;
XX
DT 30-JUN-1997 (first entry)
XX
DE Bacillus CGTase variant beta-cyclodextrin #2.
XX
KM Cyclomaltoodextrin glucanotransferase; CGTase; enzyme; Bacillus circulans;
KM thermomaneorobacter; starch; cyclomaltoodextrin; cyclodextrin; pesticide;
KM transglycosylation; cyclic glucose oligomer; glucose oligosaccharide;
KM cholesterol removal; food stabiliser; vitamin stabiliser; fungicide;
KM plastic laminate; biodegradable plastic; mutain.
XX
XX Synthetic.
OS

Key Location/Qualifiers
FH 21
FT Modified-site /label= F21Y
FT 47
FT Modified-site /label= R47X
FT /note= "X= Ala, Gln, His, Arg, Leu"
FT 87
FT Modified-site /label= I87H
FT 88
FT Modified-site /label= N88X
FT /note= "X= Pro, Asn, Lys, His"
FT 89
FT Modified-site /label= Y89X
FT /note= "X= Asp, Gly, Ala, Lys, Arg, Pro, Glu or absent"
FT 90
FT Modified-site /label= S90X
FT /note= "X= Gly, Ala"
FT 91
FT Modified-site /label= G91X
FT /note= "X= Ala, Val, Ser, Thr"
FT 92
FT Misc-difference /note= "possible insertion of Ala, Val, Gly, Phe, or Tyr"
FT 93
FT Modified-site /label= V92G
FT 94
FT Modified-site /label= N93X
FT /note= "X= Gly, His, Thr, or absent"
FT 95
FT Modified-site /label= N94X
FT /note= "X= Gln, Lys, Arg, Trp, Phe, Ser, or absent"
FT 99
FT Modified-site

FT /label= H98X
FT /note= "X= Gly, Ala"
FT 102
FT Modified-site /label= W101X
FT /note= "X= Gly, Ala"
FT 136
FT Modified-site /label= D135L
FT 141
FT Modified-site /label= H140X
FT /note= "X= Ala, Arg, Asn"
FT 146
FT Modified-site /label= S145X
FT /note= "X= Ala, Glu, Trp, Leu"
FT 147
FT Misc-difference /note= "insertion of Pro, Ala, Ile, Gln, or Ser"
FT 148
FT Modified-site /label= S146X
FT /note= "X= Pro, Ala, Phe, Gln, Ser, Trp, Ile, Arg, Glu,
FT Lys, Asp, Asn, or absent"
FT 149
FT Modified-site /label= D147X
FT /note= "X= Ala, Leu, Ile, Phe, Trp, Gly, Tyr, Arg, Asp,
FT Thr, or absent"
FT 150
FT Modified-site /label= Q148X
FT /note= "X= Gly, Asn, Ala"
FT 151
FT Modified-site /label= P149W
FT 152
FT Modified-site /label= S150A
FT 169
FT Modified-site /label= Y167X
FT /note= "X= Ala, Phe"
FT 170
FT Modified-site /label= T168S
FT 180
FT Modified-site /label= N178Y
FT 181
FT Modified-site /label= G179X
FT /note= "X= Ser, Asn, Asp"
FT 182
FT Modified-site /label= G180X
FT /note= "X= Ser, Asn, Asp"
FT 185
FT Modified-site /label= F183X
FT /note= "X= Trp, Tyr, Ala"
FT 187
FT Modified-site /label= T185X
FT /note= "X= Pro, His, Arg, Glu, Asp"
FT 194
FT Modified-site /label= K192R
FT 195
FT Modified-site /label= N193X
FT /note= "X= Gly, Ala"
FT 197
FT Modified-site /label= L195X
FT /note= "X= Leu, Ile, Trp, Phe"
FT 198
FT Modified-site /label= D196X
FT /note= "X= Ala, Gly, Asn, Ser"
FT 199
FT Modified-site /label= L197X
FT /note= "X= Asp, Glu"
FT 234
FT Modified-site /label= K232X
FT /note= "X= Gln, Leu"
FT 235
FT Modified-site /label= H233X
FT /note= "X= Gln, Asn, Ile"
FT 261
FT Modified-site /label= F259X
FT /note= "X= Phe, Trp, Ala"
FT 266
FT Modified-site

FT /label= E264Q
FT Modified-site 328
FT /label= N326X
FT /note= "X= Gln, Phe, Leu"
FT Modified-site 372
FT /label= T370N
FT Modified-site 373
FT /label= D371X
FT /note= "X= Ala, Ser, Asn, Gly"
FT Modified-site 375
FT /label= D373X
FT /note= "X= Asp, Glu, Tyr"
FT Modified-site 377
FT /label= R375X
FT /note= "X= Ala, Pro, Gly, Lys"
FT Modified-site 602
FT /label= L600X
FT /note= "X= unspecified amino acid"
XX
XX W09633267-A1.
XX
XX 24-OCT-1996.
XX
XX 22-APR-1996; 96WO-DK00179.
XX
XX 16-NOV-1995; 95DK-0001281.
XX 21-APR-1995; 95DP-0000477.
XX 17-OCT-1995; 95DK-0001173.
XX
XX (NOVO) NOVO-NORDISK AS.
XX
XX Andersen C, Dijkhuizen L, Dijkstra BW, Von Der Osten C;
XX
XX WPI: 1996-485774/48.
XX
XX New variants of cyclo:malto:dextrin glucanotransferase (CGTase) -
XX have altered substrate binding, useful for prodn. of cyclodextrin(s)
XX or linear oligosaccharide(s), opt. formed in situ in e.g. baked
XX goods
XX
XX Claim 34; Page -: 161pp; English.
XX
XX AAW17568-W17579, and AAW17592-W17605 represent mutant versions of the
XX cyclomalto-dextrin glucanotransferase (CGTase) of Bacillus circulans
XX strain 251 (see AAW06773 for wild type sequence). CGTase catalyses the
XX conversion of starch and similar substrates into cyclomalto-dextrins
XX (also known as cyclodextrins) via an intramolecular transglycosylation
XX reaction. Cyclodextrins (CD) are cyclic glucose oligomers with a
XX hydrophobic internal cavity that form inclusion complexes with many
XX small hydrophobic molecules. These CGTase mutants have a modified
XX substrate binding and/or product selectivity, compared to this sequence.
XX The mutants are created using primer mutagenesis to modify the gene
XX encoding this sequence. These sequences have greater product selectivity
XX and/or reduced product inhibition (better yields) than wild-type CGTase.
XX These mutant sequences are used to manufacture the 6 main CD types
XX (comprising 6-11 glucose units), or linear 2-12 glucose oligosaccharides,
XX optionally in situ, e.g. during production of baked goods, to stabilise
XX chemicals during their manufacture and in detergents. CD are known for
XX their usefulness in foods, e.g. as a bread-improving agent, to
XX encapsulate/stabilise/solubilise vitamins, dyes, pharmaceuticals,
XX pesticides or fungicides, to bind/remove lipophilic compounds such as
XX cholesterol (e.g. in egg yolk or butter), in plastic laminates, films
XX etc. and to make biodegradable plastics.
XX
XX Sequence 688 AA:

Query Match 37.5%; Score 3; DB 17; Length 688;
Best Local Similarity 100.0%; Pred. No. 2.2e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 ayxg 8
|||!

Db 97 ayxg 100
RESULT 20
AAW17570
ID AAW17570 standard; protein; 689 AA.
XX
XX AAW17570;
AC
XX 30-JUN-1997 (first entry)
DT
XX
XX Bacillus CGTase variant gamma-cyclodextrin #1.
DE
XX
XX Cyclomalto-dextrin glucanotransferase; CGTase; enzyme; Bacillus circulans;
KW thermobacter; starch; cyclomalto-dextrin; cyclodextrin; pesticide;
KW transglycosylation; cyclic glucose oligomer; glucose oligosaccharide;
KW cholesterol removal; food stabiliser; vitamin stabiliser; fungicide;
KW plastic laminate; biodegradable plastic; muten.
XX
XX Synthetic.
OS
XX
XX Key Location/Qualifiers
FT Modified-site 21
FT /label= F21Y
FT Modified-site 47
FT /label= R47X
FT /note= "X= Ala, Gln, His, Arg, Leu"
FT Modified-site 87
FT /label= I87H
FT Modified-site 88
FT /label= N88X
FT /note= "X= Pro, Asn, Lys, His"
FT Modified-site 89
FT /label= Y89X
FT /note= "X= Asp, Gly, Ala, Tyr, Pro or absent"
FT Modified-site 90
FT /label= S90X
FT /note= "X= Gly, Ala, Ser"
FT Modified-site 91
FT /label= G91X
FT /note= "X= Ala, Val, Ser, Gly"
FT Misc-difference 92
FT /note= "possible insertion of Ala, Val, Gly, Phe, or Tyr"
FT Modified-site 93
FT /label= V92G
FT Modified-site 94
FT /label= M93X
FT /note= "X= Gly, Asn, His, Thr, or absent"
FT Modified-site 95
FT /label= N94X
FT /note= "X= Gln, Lys, Arg, Trp, Phe, Asn, Ser, or absent"
FT Modified-site 99
FT /label= H98X
FT /note= "X= Gly, Ala"
FT Modified-site 102
FT /label= W101X
FT /note= "X= Gly, Ala, Phe, Tyr"
FT Modified-site 136
FT /label= D135L
FT Modified-site 141
FT /label= H140X
FT /note= "X= Ala, Arg, Asn"
FT Modified-site 146
FT /label= S145X
FT /note= "X= Ala, Glu, Trp, Leu"
FT Misc-difference 147
FT /note= "insertion of Pro, Ala, Ile, Gln, or Ser"
FT Modified-site 148
FT /label= S146X
FT /note= "X= Pro, Ala, Gln, Ser, Ile, Glu, or absent"
FT Modified-site 149
FT /label= D147X
FT /note= "X= Ala, Asp, Ser, Thr"

FT	Modified-site	150	/note= "Insertion of Asp, Asn, Glu, Gln, or Thr"
FT	Modified-site	151	/label= Q148X
FT	Modified-site	152	/note= "X= Asn, Gln, Asp, Glu, Arg, Lys, Phe, Tyr"
FT	Modified-site	153	/label= P149X
FT	Modified-site	154	/note= "X= Phe, Leu, Ile, Trp, Pro"
FT	Modified-site	155	/label= S150X
FT	Modified-site	156	/note= "X= Asn, Ala, Ser"
FT	Modified-site	157	/label= Y167X
FT	Modified-site	158	/note= "X= Ala, Phe, Tyr, Pro"
FT	Modified-site	159	/label= T168S
FT	Modified-site	160	/label= N178Y
FT	Modified-site	161	/label= G179X
FT	Modified-site	162	/note= "X= Gly, Ser, Asn, Asp"
FT	Modified-site	163	/label= G180X
FT	Modified-site	164	/note= "X= Gly, Ser, Asn, Asp"
FT	Modified-site	165	/label= F183X
FT	Modified-site	166	/note= "X= Phe, Trp, Tyr, Ala"
FT	Modified-site	167	/label= T185X
FT	Modified-site	168	/note= "X= Pro, His, Arg, Glu, Asp"
FT	Modified-site	169	/label= K192R
FT	Modified-site	170	/label= N193X
FT	Modified-site	171	/note= "X= Gly, Ala, Asn"
FT	Modified-site	172	/label= L195X
FT	Modified-site	173	/note= "X= Leu, Ile, Trp, Tyr, Phe"
FT	Modified-site	174	/label= D196X
FT	Modified-site	175	/note= "X= Ala, Asp, Asn, Ser"
FT	Modified-site	176	/label= L197X
FT	Modified-site	177	/note= "X= Asp, Glu, Leu"
FT	Modified-site	178	/label= K232X
FT	Modified-site	179	/note= "X= Lys, Gln, Leu"
FT	Modified-site	180	/label= H233X
FT	Modified-site	181	/note= "X= His, Gln, Asn, Ile"
FT	Modified-site	182	/label= F259X
FT	Modified-site	183	/note= "X= Phe, Trp, Tyr, Ala"
FT	Modified-site	184	/label= E264Q
FT	Modified-site	185	/label= N326X
FT	Modified-site	186	/note= "X= Gln, Phe, Leu"
FT	Modified-site	187	/label= T370N
FT	Modified-site	188	/label= D371X
FT	Modified-site	189	/note= "X= Val, Leu, Ile, Ala, Asp, Ser, Asn, Gly"
FT	Modified-site	190	/label= D373X
FT	Modified-site	191	/note= "X= Asp, Glu, Tyr"
FT	Modified-site	192	/label= R375X
FT	Modified-site	193	/note= "X= Ala, Pro, Gly, Arg, Lys"
FT	Modified-site	194	/label= L600X
FT	Modified-site	195	/note= "X= unspecified amino acid"
XX			

PN	W09633267-A1.	
XX	24-OCT-1996.	
XX	22-APR-1996;	96WO-DK00179.
XX	16-NOV-1995;	95DK-0001281.
XX	21-APR-1995;	95DK-0000477.
XX	17-OCT-1995;	95DK-0001173.
XX	(NOVO) NOVO-NORDISK AS.	
XX	Andersen C, Dijkhuizen L, Dijkstra BW, Von Der Osten C;	
XX	WPI; 1996-485774/48.	
XX	New variants of cyclo:malto:dextrin glucanotransferase (CGTase) -	
XX	have altered substrate binding, useful for prodn. of cyclodextrin(s)	
XX	or linear oligosaccharide(s), opt. formed in situ in e.g. baked	
XX	goods	
XX	Claim 31; Page -; 161pp; English.	
XX	AAW17568-W17579, and AAW17592-W17605 represent mutant versions of the	
XX	cyclomaltodextrin glucanotransferase (CGTase) of Bacillus circulans	
XX	strain 251 (see AAW06773 for wild type sequence). CGTase catalyses the	
XX	conversion of starch and similar substrates into cyclomaltodextrins	
XX	(also known as cyclodextrins) via an intramolecular transglycosylation	
XX	reaction. Cyclodextrins (CD) are cyclic glucose oligomers with a	
XX	hydrophobic internal cavity that form inclusion complexes with many	
XX	small hydrophobic molecules. These CGTase mutants have a modified	
XX	substrate binding and/or product selectivity, compared to this sequence.	
XX	The mutants are created using primer mutagenesis to modify the gene	
XX	encoding this sequence. These sequences have greater product selectivity	
XX	and/or reduced product inhibition (better yields) than wild-type CGTase.	
XX	These mutant sequences are used to manufacture the 6 main CD types	
XX	(comprising 6-11 glucose units), or linear 2-12 glucose oligosaccharides,	
XX	optionally in situ, e.g. during production of baked goods, to stabilise	
XX	chemicals during their manufacture and in detergents. CD are known for	
XX	their usefulness in foods, e.g. as a bread-improving agent, to	
XX	encapsulate/stabilise/solubilise vitamins, dyes, pharmaceuticals,	
XX	pesticides or fungicides, to bind/remove lipophilic compounds such as	
XX	cholesterol (e.g. in egg yolk or butter), in plastic laminates, films	
XX	etc. and to make biodegradable plastics.	
SQ	Sequence 689 AA:	
Query Match 37.5%; Score 3; DB 17; Length 689;		
Best Local Similarity 100.0%; Pred. No. 2.2e+03;		
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
OY	5 ayxg 8	
DB	97 ayxg 100	
RESULT 21		
AAW17574		
ID	AAW17574 standard; protein; 689 AA.	
XX	AAW17574;	
XX	30-JUN-1997 (first entry)	
XX	Bacillus CGTase variant gamma-cyclodextrin #2.	
XX	Cyclomaltodextrin glucanotransferase; CGTase; enzyme; Bacillus circulans;	
XX	thermoanaerobacter; starch; cyclomaltodextrin; cyclodextrin; pesticide;	
XX	transglycosylation; cyclic glucose oligomer; glucose oligosaccharide;	
XX	cholesterol removal; food stabiliser; vitamin stabiliser; fungicide;	
XX	plastic laminate; biodegradable plastic; muteln.	

```

OS      Synthetic.
XX      Key
FH      Modified-site
FT      /location/Qualifiers
FT      21
FT      /label= F21Y
FT      47
FT      /label= R47X
FT      /note= "X= Ala, Gln, His, Arg, Leu"
FT      87
FT      /label= I87H
FT      88
FT      /label= N88X
FT      /note= "X= Pro, Asn, Lys, His"
FT      89
FT      /label= Y89X
FT      /note= "X= Asp, Gly, Ala, Pro or absent"
FT      90
FT      /label= S90X
FT      /note= "X= Gly, Ala"
FT      91
FT      /label= G91X
FT      /note= "X= Ala, Val, Ser"
FT      92
FT      /note= "possible insertion of Ala, Val, Gly, Phe, or Tyr"
FT      93
FT      /label= V92G
FT      94
FT      /label= N93X
FT      /note= "X= Gly, His, Thr, or absent"
FT      95
FT      /label= N94X
FT      /note= "X= Gln, Lys, Arg, Trp, Phe, Ser, or absent"
FT      99
FT      /label= H98X
FT      /note= "X= Gly, Ala"
FT      102
FT      /label= W101X
FT      /note= "X= Gly, Ala, Phe, Tyr"
FT      136
FT      /label= D135L
FT      141
FT      /label= H140X
FT      /note= "X= Ala, Arg, Asn"
FT      146
FT      /label= S145X
FT      /note= "X= Ala, Glu, Trp, Leu"
FT      147
FT      /note= "insertion of Pro, Ala, Ile, Gln, or Ser"
FT      148
FT      /label= S146X
FT      /note= "X= Pro, Ala, Gln, Ser, Ile, Glu"
FT      149
FT      /label= D147X
FT      /note= "X= Ala, Asp, Ser, Thr"
FT      150
FT      /note= "insertion of Asp, Asn, Glu, Gln, or Thr"
FT      151
FT      /label= Q148X
FT      /note= "X= Asn, Asp, Glu, Arg, Lys, Phe, Tyr"
FT      152
FT      /label= P149X
FT      /note= "X= Phe, Leu, Ile, Trp"
FT      153
FT      /label= S150A
FT      /note= "X= Asn, Ala, Ser"
FT      170
FT      /label= Y167X
FT      /note= "X= Ala, Phe, Pro"
FT      171
FT      /label= T168S
FT      181
FT      /label= N178Y
FT      182

```

```

FT      /label= G179X
FT      /note= "X= Ser, Asn, Asp"
FT      183
FT      /label= G180X
FT      /note= "X= Ser, Asn, Asp"
FT      186
FT      /label= F183X
FT      /note= "X= Trp, Tyr, Ala"
FT      188
FT      /label= T185X
FT      /note= "X= Pro, His, Arg, Glu, Asp"
FT      195
FT      /label= K192R
FT      196
FT      /label= N193X
FT      /note= "X= Gly, Ala"
FT      198
FT      /label= L195X
FT      /note= "X= Leu, Ile, Trp, Phe"
FT      199
FT      /label= D196X
FT      /note= "X= Ala, Gly, Asn, Ser"
FT      200
FT      /label= L197X
FT      /note= "X= Asp, Glu"
FT      235
FT      /label= K232X
FT      /note= "X= Gln, Leu"
FT      236
FT      /label= H233X
FT      /note= "X= Gln, Asn, Ile"
FT      262
FT      /label= F259X
FT      /note= "X= Phe, Trp, Ala"
FT      267
FT      /label= E264Q
FT      329
FT      /label= N326X
FT      /note= "X= Gln, Phe, Leu"
FT      373
FT      /label= T370N
FT      374
FT      /label= D371X
FT      /note= "X= Val, Leu, Ile, Ala, Ser, Asn, Gly"
FT      376
FT      /label= D373X
FT      /note= "X= Asp, Glu, Tyr"
FT      378
FT      /label= R375X
FT      /note= "X= Ala, Pro, Gly, Lys"
FT      603
FT      /label= L600X
FT      /note= "X= unspecified amino acid"

WO9633267-A1.
PD      24-OCT-1996.
XX      22-APR-1996; 96WO-DK00179.
XX      16-NOV-1995; 95DK-0001281.
XX      21-APR-1995; 95DK-0000477.
XX      17-OCT-1995; 95DK-0001173.
XX      (NOVO ) NOVO-NORDISK AS.
XX      Andersen C, Dijkhuizen L, Dijkstra BW, Von Der Osten C;
XX      WPI; 1996-485774/48.
XX      New variants of cyclo:malto:dextrin glucanotransferase (CGTase) -
XX      have altered substrate binding, useful for produ. of cyclodextrin(s)
XX      or linear oligosaccharide(s), opt. formed in situ in e.g. baked

```

PT goods
XX
PS Claim 34; Page -: 161pp; English.
XX
CC AAM17568-W17579, and AAM17592-W17605 represent mutant versions of the
CC cyclomaltopectin glucanotransferase (CGTase) of *Bacillus circulans* the
CC strain 231 (see AAM06773 for wild type sequence). CGTase catalyses the
CC conversion of starch and similar substrates into cyclomaltopectins
CC (also known as cyclodextrins) via an intramolecular transglycosylation
CC reaction. Cyclodextrins (CD) are cyclic glucose oligomers with a
CC hydrophobic internal cavity that form inclusion complexes with many
CC small hydrophobic molecules. These CGTase mutants have a modified
CC substrate binding and/or product selectivity, compared to this sequence.
CC The mutants are created using primer mutagenesis to modify the gene
CC encoding this sequence. These sequences have greater product selectivity
CC and/or reduced product inhibition (better yields) than wild-type CGTase.
CC These mutant sequences are used to manufacture the 6 main CD types
CC (comprising 6-11 glucose units), or linear 2-12 glucose oligosaccharides,
CC optionally in situ, e.g. during production of baked goods, to stabilise
CC chemicals during their manufacture and in detergents. CD are known for
CC their usefulness in foods, e.g. as a bread-improving agent, to
CC encapsulate/stabilise/solubilise vitamins, dyes, pharmaceuticals,
CC pesticides or fungicides, to bind/remove lipophilic compounds such as
CC cholesterol (e.g. in egg yolk or butter), in plastic laminates, films
CC etc. and to make biodegradable plastics.
XX
SQ Sequence 689 AA;

Query Match 37.5%; Score 3; DB 17; Length 689;
Best Local Similarity 100.0%; Pred. No. 2.2e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ayxg 8
|||
Db 97 ayxg 100

RESULT 22

ID AAM17578
AAM17578 standard; protein; 689 AA.

XX AAM17578;

XX 30-JUN-1997 (first entry)

XX *Bacillus* CGTase variant gamma-cyclodextrin #3.

XX Cyclomaltopectin glucanotransferase; CGTase; enzyme; *Bacillus circulans*;
KW thermannorobacter; starch; cyclomaltopectin; cyclodextrin; pesticide;
KW transglycosylation; cyclic glucose oligomer; glucose oligosaccharide;
KW cholesterol removal; food stabiliser; vitamin stabiliser; fungicide;
KW plastic laminate; biodegradable plastic; mucin.
XX
OS Synthetic.

XX
FH Key location/Qualifiers
FT Modified-site 21 /label= F21Y
FT Modified-site 47 /label= R47X
FT Modified-site 87 /note= "X= Ala, Gln, His, Arg, Leu"
FT Modified-site 88 /label= 187H
FT Modified-site 89 /note= "X= Asn, Lys, His"
FT Modified-site 90 /label= Y89X
FT Modified-site 90 /note= "X= Gly, Ala, Tyr, Pro or absent"
FT Modified-site 90 /label= S90X
FT /note= "X= Gly, Ala"

FT Modified-site 91 /label= G91X
FT /note= "X= Ala, Val, Ser, Gly"
FT Misc-difference 92 /note= "possible insertion of Ala, Val, Gly, or Tyr"
FT Modified-site 94 /label= N93X
FT /note= "X= Asn, His, Thr, or absent"
FT Modified-site 95 /label= N94X
FT /note= "X= Gln, Lys, Arg, Trp, Phe, Asn, or absent"
FT Modified-site 99 /label= H98X
FT /note= "X= Gly, Ala"
FT Modified-site 102 /label= W101X
FT /note= "X= Gly, Ala, Phe, Tyr"
FT Modified-site 136 /label= D135L
FT /label= H140X
FT /label= H140X
FT /note= "X= Ala, Arg, Asn"
FT Modified-site 146 /label= S145X
FT /note= "X= Ala, Glu, Trp, Leu"
FT Misc-difference 147 /note= "insertion of Pro, Ala, Ile, Gln, or Ser"
FT Modified-site 148 /label= S146X
FT /note= "X= Pro, Ala, Gln, Ser, Ile, or absent"
FT Modified-site 149 /label= D147X
FT /note= "X= Ala, Asp, Ser"
FT Modified-site 150 /note= "X= Ala, Phe, Leu, Ile, Trp"
FT /note= "insertion of Asp, Asn, Glu, Gln, or Thr"
FT Modified-site 151 /label= Q148X
FT /note= "X= Asn, Gln, Glu, Arg, Lys, Phe, Tyr"
FT Modified-site 152 /label= P149X
FT /note= "X= Phe, Leu, Ile, Trp"
FT Modified-site 153 /label= S150X
FT /note= "X= Asn, Ala, Ser"
FT Modified-site 170 /label= Y167X
FT /note= "X= Ala, Phe, Pro"
FT Modified-site 171 /label= T168S
FT /label= G179X
FT /note= "X= Ser, Asn, Asp"
FT Modified-site 183 /label= G180X
FT /note= "X= Ser, Asn, Asp"
FT Modified-site 186 /label= F183X
FT /note= "X= Trp, Tyr, Ala"
FT Modified-site 188 /label= T185X
FT /note= "X= Pro, His, Arg, Glu, Asp"
FT Modified-site 196 /label= N193X
FT /note= "X= Gly, Ala"
FT Modified-site 198 /label= L195X
FT /note= "X= Leu, Ile, Trp, Tyr"
FT Modified-site 199 /label= D196X
FT /note= "X= Ala, Gly, Asn, Ser"
FT Modified-site 200 /label= L197X
FT /note= "X= Asp, Glu"

FT Modified-site 235
 FT /label= K232X
 FT /note= "X= Gln, Leu"
 FT Modified-site 236
 FT /label= H233X
 FT /note= "X= Gln, Asn, Ile"
 FT Modified-site 262
 FT /label= F259X
 FT /note= "X= Phe, Trp, Ala"
 FT Modified-site 267
 FT /label= E264Q
 FT Modified-site 329
 FT /label= N326X
 FT /note= "X= Gln, Phe, Leu"
 FT Modified-site 373
 FT /label= T370N
 FT Modified-site 374
 FT /label= D371X
 FT /note= "X= Val, Leu, Ile, Ala, Ser, Asn, Gly"
 FT Modified-site 376
 FT /label= D373X
 FT /note= "X= Asp, Glu"
 FT Modified-site 378
 FT /label= R375X
 FT /note= "X= Ala, Pro, Gly, Lys"
 FT Modified-site 603
 FT /label= L600X
 FT /note= "X= unspecified amino acid"
 FT
 PN W09633267-A1.
 XX
 PD 24-OCT-1996.
 XX
 PF 22-APR-1996; 96WO-DK00179.
 XX
 PR 16-NOV-1995; 95DK-0001281.
 PR 21-APR-1995; 95DK-0000477.
 PR 17-OCT-1995; 95DK-0001173.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Andersen C, Dijkhuizen L, Dijkstra BW, Von Der Osten C;
 XX
 DR WPI: 1996-485774/48.
 XX
 PT New variants of cyclo:malto:dextrin glucanotransferase (CGTase) -
 PT have altered substrate binding, useful for prodn. of cyclodextrin(s)
 PT or linear oligosaccharide(s), opt. formed in situ in e.g. baked
 PT goods
 XX
 PS Claim 40; Page -: 161pp; English.
 XX
 CC AAM17568-W17579, and AAM17592-W17605 represent mutant versions of the
 CC cyclomaltodextrin glucanotransferase (CGTase) of *Bacillus circulans*
 CC strain 251 (see AAM06773 for wild type sequence). CGTase catalyses the
 CC conversion of starch and similar substrates into cyclomaltodextrins
 CC (also known as cyclodextrins) via an intramolecular transglycosylation
 CC reaction. Cyclodextrins (CD) are cyclic glucose oligomers with a
 CC hydrophobic internal cavity that form inclusion complexes with many
 CC small hydrophobic molecules. These CGTase mutants have a modified
 CC substrate binding and/or product selectivity, compared to this sequence.
 CC The mutants are created using primer mutagenesis to modify the gene
 CC encoding this sequence. These sequences have greater product selectivity
 CC and/or reduced product inhibition (better yields) than wild-type CGTase.
 CC These mutant sequences are used to manufacture the 6 main CD types
 CC (comprising 6-11 glucose units), or linear 2-12 glucose oligosaccharides,
 CC optionally in situ, e.g. during production of baked goods, to stabilise
 CC chemicals during their manufacture and in detergents. CD are known for
 CC their usefulness in foods, e.g. as a bread-improving agent, to
 CC encapsulate/stabilise/solubilise vitamins, dyes, pharmaceuticals,
 CC pesticides or fungicides, to bind/remove lipophilic compounds such as
 CC cholesterol (e.g. in egg yolk or butter), in plastic laminates, films
 CC etc. and to make biodegradable plastics.

XX
 SQ Sequence 689 AA;
 Query Match 37.5%; Score 3; DB 17; Length 689;
 Best Local Similarity 100.0%; Pred. No. 2.2e+03;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 5 ayxg 8
 ||||
 DB 97 ayxg 100
 RESULT 23
 AAR42556
 ID AAR42556 standard; peptide; 3 AA.
 XX
 AC AAR42556;
 XX
 DT 07-DEC-1994 (first entry)
 XX
 DE ACE inhibitor SP3.
 XX
 KW Angiotensin converting enzyme; ACE; inhibitor; SP5; SP4; SP3;
 KW hypertension; blood pressure.
 XX
 OS Synthetic.
 XX
 PN JP06065288-A.
 XX
 PD 08-MAR-1994.
 XX
 PF 19-AUG-1992; 92JP-0220270.
 XX
 PR 19-AUG-1992; 92JP-0220270.
 XX
 PA (APIA-) API KK.
 XX
 DR WPI: 1994-115194/14.
 XX
 PT New tri-, tetra- and penta-peptide(s), e.g. Trp-Lys-Tyr - are ACE
 PT inhibitors useful for treatment or prophylaxis of hypertension
 XX
 PS Claim 1; Page 2; 5pp; Japanese.
 XX
 CC Peptides SP3, SP4 and SP5 have ACE inhibiting activity. They can be
 CC prep'd. easily and in high yield. They are useful for treatment or
 CC prophylaxis of hypertension.
 XX
 SQ Sequence 3 AA;
 Query Match 25.0%; Score 2; DB 15; Length 3;
 Best Local Similarity 100.0%; Pred. No. 4.3e+05;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 wk 2
 ||
 DB 1 wk 2
 RESULT 24
 AAP10370
 ID AAP10370 standard; peptide; 4 AA.
 XX
 AC AAP10370;
 XX
 DT 16-DEC-1992 (first entry)
 XX
 DE Generic enkephalin-like analgesic peptide.
 XX
 KW painkiller.
 XX

```

OS      Synthetic.
XX      Key
FH      Location/Qualifiers
FT      MISC-difference 1
FT      /label= OTHER
FT      /note= "N-methyl- N-ethyl-, N-n-propyl"
FT      MISC-difference 2
FT      /label= Ala, Abu, Nva, Val, Nle, Leu, Ile, Met, Ser, Thr
FT      /note= "allylglycine, cyclopropylmethylglycine,
FT      S-methylcysteine, methioninesulphoxide,
FT      S-methylcysteine sulphoxide, homoserine, or
FT      Ala substituted with C1-C4 primary or
FT      secondary alkyl, or C1-C2 hydroxyalkyl.
FT      NOTE: all amino acids are D- form."
FT      MISC-difference 4
FT      /label= OTHER
FT      /note= "p-fluoro-L-Phe, substituted at the amino N with
FT      N-methyl, N-ethyl, N-n-propyl, N-isopropyl,
FT      N-n-butyl, N-isobutyl, N-sec-butyl
FT      N-cyclopropylmethyl, N-allyl, or N-propargyl
FT      The residue may be derivatised to either its
FT      amide primary alcohol, or C1-C3 ester derivative"
PN      US4265808-A.
XX
PD      05-MAY-1981.
XX
PF      17-DEC-1979; 79US-0104529.
XX
PR      17-DEC-1979; 79US-0104529.
XX
PA      (ELIL ) ELI LILLY & CO.
XX
PI      Gesellschaft PD, Shuman RT;
XX
PS      Tetra:peptide(s) - useful as analgesic and neuroleptic agents
XX
XX      Disclosure; Column 4; 13pp; English.
XX
CC      This generic sequence is representative of a class of enkephalin
CC      analogues having a high level of analgesic activity. They are
CC      halogenated tetrapeptides which are structurally highly specific in
CC      terms both of the identity and position of the halogen. The
CC      peptides have p-fluoro-substituted L-phenylalanine at position 4.
XX
SO      Sequence 4 AA:

Query Match          25.0%; Score 2; DB 2; Length 4;
Best Local Similarity 100.0%; Pred. No. 4.3e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      6 yxg 8
        |||
DB      1 yxg 3

RESULT 25
AAP10372
ID      AAP10372 standard; ProteIn; 4 AA.
XX
AC      AAP10372;
XX
DT      16-DEC-1992 (first entry)
XX
XX      Enkephalin-like analgesic peptide #2.
XX
XX      painkiller.
XX
XX      Synthetic.
XX

```

```

FH Key Location/Qualifiers
FT Misc-difference 1 /note= "N-methyl-"
FT FT Misc-difference 2 /label= Abu
FT FT /note= "D- form"
FT Misc-difference 4
FT FT /label= OTHER
FT FT /note= "N-ethyl-p-fluoro-L-Phe, amidated"
XX
XX US4265808-A.
XX PD 05-MAY-1981.
XX 17-DEC-1979; 79US-0104529.
XX PF 17-DEC-1979; 79US-0104529.
XX PR 17-DEC-1979; 79US-0104529.
XX
XX (ELIL ) ELI LILLY & CO.
XX PA
XX PI Gesellchen PD, Shuman RT;
XX
XX WPI; 1981-38016D/21 (38016D).
XX
XX Tetra:peptide(s) - useful as analgesic and neuroleptic agents
PT
PS Disclosure; Column 5; 13pp: English.
XX
XX This generic sequence is representative of a class of enkephalin
CC analogues having a high level of analgesic activity. They are
CC halogenated tetrapeptides which are structurally highly specific in
CC terms both of the identity and position of the halogen. The
CC peptides have p-fluoro-substituted L-phenylalanine at position 4.
XX
XX Sequence 4 AA:
SQ
Query Match 25.0%; Score 2; DB 2; Length 4;
Best Local Similarity 100.0%; Pred. No. 4.3e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 yxg 8
| | |
Db 1 yxg 3

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:57:30 ; Search time 41.59 Seconds
(without alignments)
4.329 Million cell updates/sec

Title: 09-185908-1B
Perfect score: 8
Sequence: 1 wxxxayxg 8

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 212252 seqs, 22503292 residues

Word size : 0
Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2	25.0	3	4	US-09-461-697-405
2	2	25.0	4	1	US-08-079-445-3
3	2	25.0	4	1	US-07-840-077A-7
4	2	25.0	4	1	US-08-176-938-20
5	2	25.0	4	1	US-08-428-488-59
6	2	25.0	4	1	US-08-428-488-60
7	2	25.0	4	1	US-08-428-488-61
8	2	25.0	4	1	US-08-428-488-62
9	2	25.0	4	1	US-08-428-488-71
10	2	25.0	4	1	US-08-428-488-72
11	2	25.0	4	1	US-08-428-488-73
12	2	25.0	4	1	US-08-428-488-74
13	2	25.0	4	1	US-08-487-006-150
14	2	25.0	4	1	US-08-487-006-151
15	2	25.0	4	1	US-08-487-006-152
16	2	25.0	4	1	US-08-487-006-153
17	2	25.0	4	1	US-08-454-950-7
18	2	25.0	4	1	US-08-434-761-3
19	2	25.0	4	1	US-08-338-890B-1
20	2	25.0	4	1	US-07-923-260A-11
21	2	25.0	4	1	US-08-654-949-7
22	2	25.0	4	1	US-08-798-897-49
23	2	25.0	4	1	US-08-463-224-2
24	2	25.0	4	2	US-08-463-377-2
25	2	25.0	4	2	US-08-441-871-63
26	2	25.0	4	2	US-08-441-871-66
27	2	25.0	4	2	US-08-978-523-49

28	2	25.0	4	2	US-08-488-659A-150	Sequence 150, App
29	2	25.0	4	2	US-08-488-659A-151	Sequence 151, App
30	2	25.0	4	2	US-08-488-659A-152	Sequence 152, App
31	2	25.0	4	2	US-08-488-659A-153	Sequence 153, App
32	2	25.0	4	2	US-08-488-659A-153	Sequence 153, App
33	2	25.0	4	2	US-08-488-659A-153	Sequence 153, App
34	2	25.0	4	2	US-08-340-208B-1	Sequence 1, Appl1
35	2	25.0	4	2	US-08-484-905-124	Sequence 124, App
36	2	25.0	4	2	US-08-651-179B-2	Sequence 2, Appl1
37	2	25.0	4	2	US-08-651-179B-3	Sequence 3, Appl1
38	2	25.0	4	2	US-08-997-263-1	Sequence 1, Appl1
39	2	25.0	4	3	US-08-481-985B-124	Sequence 124, App
40	2	25.0	4	3	US-08-912-272-8	Sequence 8, Appl1
41	2	25.0	4	3	US-09-315-861-5	Sequence 5, Appl1
42	2	25.0	4	3	US-08-981-122-25	Sequence 25, Appl1
43	2	25.0	4	4	US-08-435-568A-19	Sequence 19, Appl1
44	2	25.0	4	4	US-08-370-476-124	Sequence 124, App
45	2	25.0	4	4	US-09-051-986-26	Sequence 26, Appl1
46	2	25.0	4	4	US-08-682-767-29	Sequence 29, Appl1
47	2	25.0	4	4	US-08-682-767-30	Sequence 30, Appl1
48	2	25.0	4	6	5187077-30	Patent No. 5187077
49	2	25.0	4	6	5427925-28	Patent No. 5427925
50	2	25.0	5	1	US-07-657-769B-36	Sequence 36, Appl1
					US-07-630-163B-18	Sequence 18, Appl1

ALIGNMENTS

RESULT 1
US-09-461-697-405
; Sequence 405, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: LO, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Purnam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ. ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 405
; LENGTH: 3
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-461-697-405

Query Match 25.0%; Score 2; DB 4; Length 3;
Best local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wk 2
11
DB 2 wk 3

RESULT 2
US-08-079-445-3
; Sequence 3, Application US/08079445
; Patent No. 5440016
; GENERAL INFORMATION:
; APPLICANT: Blondelle, Sylvie E.
; APPLICANT: Pinilla, Clemencia
; APPLICANT: Eichler, Jutta
; APPLICANT: Houghten, Richard A.

;; TITLE OF INVENTION: PEPTIDES HAVING ANTI-MICROBIAL,
;; TITLE OF INVENTION: HEMOLYTIC, ANTI-MELITTIN OR ANTI-ENZYMATIC ACTIVITY
;; NUMBER OF SEQUENCES: 15
;; CORRESPONDENCE ADDRESS:
;; ADDRESS: CAMPBELL AND FLORES
;; STREET: 4370 La Jolla Village Drive, Suite 700
;; CITY: San Diego
;; STATE: California
;; COUNTRY: United States of America
;; ZIP: 92122
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/079,445
;; FILING DATE: 18-JUN-1993
;;
;; CLASSIFICATION: 514
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Campbell, Cathryn A.
;; REGISTRATION NUMBER: 31,815
;; REFERENCE/DOCKET NUMBER: P-HP 9648
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (619) 535-9001
;; TELEFAX: (619) 535-8949
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 4 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;;
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 1
;; OTHER INFORMATION: /label= Xaa
;; OTHER INFORMATION: /note= "Xaa-KFmoc"
;;
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 4
;; OTHER INFORMATION: /label= Xaa
;; OTHER INFORMATION: /note= "Xaa-any amino acid"
;;
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 4
;; OTHER INFORMATION: /note= "C-terminal amino acid is
;; OTHER INFORMATION: amended"
;;
;; US-08-079-445-3
;;
Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;;
QY 1 wk 2
11
Db 2 wk 3
;;
RESULT 3
US-07-840-077A-7
; Sequence 7, Application US/07840077A
; Patent No. 5443816
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT: Zamora, Paul O.
; APPLICANT: Rhodes, Buck A.
; TITLE OF INVENTION: Peptide-Metal Ion
; TITLE OF INVENTION: Pharmaceutical Preparation and Method
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESS: Rhomed Incorporated
; STREET: 4261 Balloon Park
; CITY: Albuquerque

;; STATE: NM
;; COUNTRY: U.S.A.
;; ZIP: 87109-5802
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb
;; MEDIUM TYPE: Storage
;; COMPUTER: IBM PC/XT/AT, IBM PS/2 or
;; COMPUTER: compatibles
;; OPERATING SYSTEM: PC-DOS or MS-DOS
;; SOFTWARE: Wordperfect 6.0a for Windows
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/840,077A
;; FILING DATE: 20-FEB-1992
;;
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 07/816,477
;; FILING DATE: 03-JAN-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Deborah A. Peacock
;; REGISTRATION NUMBER: 31,649
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (505) 242-9677
;; TELEFAX: (505) 243-2542
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 4 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: Peptide
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;;
;; US-07-840-077A-7

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;;
QY 1 wk 2
11
Db 2 wk 3
;;
RESULT 4
US-08-176-938-20
; Sequence 20, Application US/08176938
; Patent No. 5602099
; GENERAL INFORMATION:
; APPLICANT: Schiller, Peter W.
; TITLE OF INVENTION: New Peptides
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESS: White & Case
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/176,938
; FILING DATE: 04-JAN-1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Steiner Ph.D., Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-080
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-819-8783
TELEFAX: 212-354-8113
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 2
OTHER INFORMATION: /label= "Tic
OTHER INFORMATION: /note= "1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 3
OTHER INFORMATION: /label= "Phg
OTHER INFORMATION: /note= "phenylglycine"
US-08-176-938-20

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 yxg 8
111
Db 1 YXG 3

RESULT 5
US-08-428-488-59
Sequence 59, Application US/08428488
Patent No. 5624894
GENERAL INFORMATION:
APPLICANT: BODOR, Nicholas S.
TITLE OF INVENTION: BRAIN-ENHANCED DELIVERY OF NEUROACTIVE
TITLE OF INVENTION: PEPTIDES BY SEQUENTIAL METABOLISM
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/428,488
FILING DATE: 27-APR-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Baumeister, Mary Katherine
REGISTRATION NUMBER: 26,254
REFERENCE/DOCKET NUMBER: 028724-087
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /note= "Position 1 = nicotinoyl-Ala."

FEATURE:
NAME/KEY: Modified-site
LOCATION: 3
OTHER INFORMATION: /note= "Position 3 = Tyr (O-tBu)."
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /note= "Position 4 = Arg (jPmc)-Cholesteryl
US-08-428-488-59

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6
111
Db 2 AY 3

RESULT 6
US-08-428-488-60
Sequence 60, Application US/08428488
Patent No. 5624894
GENERAL INFORMATION:
APPLICANT: BODOR, Nicholas S.
TITLE OF INVENTION: BRAIN-ENHANCED DELIVERY OF NEUROACTIVE
TITLE OF INVENTION: PEPTIDES BY SEQUENTIAL METABOLISM
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/428,488
FILING DATE: 27-APR-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Baumeister, Mary Katherine
REGISTRATION NUMBER: 26,254
REFERENCE/DOCKET NUMBER: 028724-087
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /note= "Position 1 = Trigonoethyl-Ala."
FEATURE:
NAME/KEY: Modified-site
LOCATION: 3
OTHER INFORMATION: /note= "Position 3 = Tyr (O-tBu)."
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /note= "Position 4 = Arg (jPmc)-Cholesteryl
OTHER INFORMATION: ester."

US-08-428-488-60

Query Match 25.0%; Score 2; DB 1; Length 4;

Best Local Similarity 100.0%; Pred. No. 1.6e+05; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6
11
Db 2 AY 3RESULT 7
US-08-428-488-61

; Sequence 61, Application US/08428488

; Patent No. 5624894

; GENERAL INFORMATION:

; APPLICANT: BODOR, Nicholas S.

; TITLE OF INVENTION: BRAIN-ENHANCED DELIVERY OF NEUROACTIVE

; NUMBER OF SEQUENCES: 107

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Burns, Doane, Swecker & Mathis

; STREET: P.O. Box 1404

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: United States

; ZIP: 22133-1404

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/428,488

; FILING DATE: 27-Apr-1995

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Baumeister, Mary Katherine

; REGISTRATION NUMBER: 26,254

; REFERENCE/DOCKET NUMBER: 028724-087

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 836-6620

; INFORMATION FOR SEQ ID NO: 61:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; FEATURE:

; NAME/KEY: Modified-site

; LOCATION: 1

; OTHER INFORMATION: /note= "Position 1 = Trigonellyl-Ala"

; FEATURE:

; NAME/KEY: Modified-site

; LOCATION: 4

; OTHER INFORMATION: /note= "Position 4 = Arg-Cholesteryl ester."

; US-08-428-488-61

OY 5 ay 6
11
Db 2 AY 3OY 5 ay 6
11
Db 2 AY 3

RESULT 8

US-08-428-488-62

; Sequence 62, Application US/08428488

; Patent No. 5624894

; GENERAL INFORMATION:

; APPLICANT: BODOR, Nicholas S.

; TITLE OF INVENTION: BRAIN-ENHANCED DELIVERY OF NEUROACTIVE

; NUMBER OF SEQUENCES: 107

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Burns, Doane, Swecker & Mathis

; STREET: P.O. Box 1404

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: United States

; ZIP: 22133-1404

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/428,488

; FILING DATE: 27-Apr-1995

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Baumeister, Mary Katherine

; REGISTRATION NUMBER: 26,254

; REFERENCE/DOCKET NUMBER: 028724-087

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 836-6620

; INFORMATION FOR SEQ ID NO: 62:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; FEATURE:

; NAME/KEY: Modified-site

; LOCATION: 1

; OTHER INFORMATION: /note= "Position 1 = 1,4-Dihydrotrigonellyl-Ala."

; FEATURE:

; NAME/KEY: Modified-site

; LOCATION: 4

; OTHER INFORMATION: /note= "Position 4 = Arg-Cholesteryl ester."

; US-08-428-488-62

OY 5 ay 6
11
Db 2 AY 3OY 5 ay 6
11
Db 2 AY 3

RESULT 9

US-08-428-488-71

; Sequence 71, Application US/08428488

; Patent No. 5624894

; GENERAL INFORMATION:

; APPLICANT: BODOR, Nicholas S.

; TITLE OF INVENTION: BRAIN-ENHANCED DELIVERY OF NEUROACTIVE

; NUMBER OF SEQUENCES: 107

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Burns, Doane, Swecker & Mathis

; STREET: P.O. Box 1404

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: United States

; ZIP: 22133-1404

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/428,488
FILING DATE: 27-APR-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Baumeister, Mary Katherine
REGISTRATION NUMBER: 26,254
REFERENCE/DOCKET NUMBER: 028724-087
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 71:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /note= "Position 1 = nicotinoyl-Pro."
FEATURE:
NAME/KEY: Modified-site
LOCATION: 3
OTHER INFORMATION: /note= "Position 3 = Tyr (O-tbu)."
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /note= "Position 4 = Arg (j-Pmc)-cholesteryl"
OTHER INFORMATION: ester."
US-08-428-488-71

Query Match      25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ay 6
DB 2 ay 3

RESULT 10
US-08-428-488-72
Sequence 72, Application US/08428488
Patent No. 5624894
GENERAL INFORMATION:
APPLICANT: BODOR, Nicholas S.
TITLE OF INVENTION: BRAIN-ENHANCED DELIVERY OF NEUROACTIVE
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/428,488
FILING DATE: 27-APR-1995
CLASSIFICATION: 514

```

```

ATTORNEY/AGENT INFORMATION:
NAME: Baumeister, Mary Katherine
REGISTRATION NUMBER: 26,254
REFERENCE/DOCKET NUMBER: 028724-087
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 72:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /note= "Position 1 = trigonellyl-Pro"
OTHER INFORMATION: methylsulfate."
FEATURE:
NAME/KEY: Modified-site
LOCATION: 2
OTHER INFORMATION: /note= "Position 2 = Tyr (O-tbu)."
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /note= "Position 4 = Arg (j-Pmc)-cholesteryl"
OTHER INFORMATION: ester."
US-08-428-488-72

Query Match      25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ay 6
DB 2 ay 3

RESULT 11
US-08-428-488-73
Sequence 73, Application US/08428488
Patent No. 5624894
GENERAL INFORMATION:
APPLICANT: BODOR, Nicholas S.
TITLE OF INVENTION: BRAIN-ENHANCED DELIVERY OF NEUROACTIVE
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/428,488
FILING DATE: 27-APR-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Baumeister, Mary Katherine
REGISTRATION NUMBER: 26,254
REFERENCE/DOCKET NUMBER: 028724-087
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 73:

```

SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /note="Position 1 = trigonellyl-pro
OTHER INFORMATION: methylsulfate."
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /note="Position 4 = Arg-Cholesteryl
OTHER INFORMATION: ester."
US-08-428-488-73

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6
11
DB 2 AY 3

RESULT 12
US-08-428-488-74
Sequence 74, Application US/08428488
Patent No. 5624894
GENERAL INFORMATION:
APPLICANT: BODOR, Nicholas S.
TITLE OF INVENTION: BRAIN-ENHANCED DELIVERY OF NEUROACTIVE
TITLE OF INVENTION: PEPTIDES BY SEQUENTIAL METABOLISM
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P. O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/428,488
FILING DATE: 27-APR-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Baumeister, Mary Katherine
REGISTRATION NUMBER: 26,254
REFERENCE/DOCKET NUMBER: 028724-087
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /note="Position 1 = 1,4-dihydrotrigonellyl-pro."
FEATURE:
NAME/KEY: Modified-site

LOCATION: 4
OTHER INFORMATION: /note="Position 4 = Arg-Cholesteryl
OTHER INFORMATION: ester."
US-08-428-488-74

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ay 6
11
DB 2 AY 3

RESULT 13
US-08-487-006-150
Sequence 150, Application US/08487006
Patent No. 5641861
GENERAL INFORMATION:
APPLICANT: Dooley, Colette T.
TITLE OF INVENTION: No. 5641861 Mu Opioid Receptor Ligands:
NUMBER OF SEQUENCES: 222
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,006
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-TP 1706
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 150:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
TOPOLOGY: linear
FEATURE:
NAME/KEY: Peptide
LOCATION: 2
OTHER INFORMATION: /note="Xaa is (D)Nve"
FEATURE:
NAME/KEY: Peptide
LOCATION: 4
OTHER INFORMATION: /note="Xaa is Nap."
FEATURE:
NAME/KEY: Peptide
LOCATION: 4
OTHER INFORMATION: /note="Amino acid is amidated at
the C-terminal."
US-08-487-006-150

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 YXG 8
111
DB 1 YXG 3

RESULT 14
US-08-487-006-151
Sequence 151, Application US/08487006
Patent No. 5641861
GENERAL INFORMATION:
APPLICANT: Dooley, Colette T.
APPLICANT: Houghten, Richard A.
TITLE OF INVENTION: No. 5641861el Mu Opioid Receptor Ligands:
NUMBER OF SEQUENCES: 222
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,006
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-TP 1706
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 151:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
TOPOLOGY: linear
FEATURE:
NAME/KEY: Peptide
LOCATION: 2
OTHER INFORMATION: /note= "Xaa is (D)Nle."
FEATURE:
NAME/KEY: Peptide
LOCATION: 4
OTHER INFORMATION: /note= "Xaa is Nap."
FEATURE:
NAME/KEY: Peptide
LOCATION: 4
OTHER INFORMATION: /note= "Amino acid is amidated at
OTHER INFORMATION: the C-terminal."
US-08-487-006-151

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 YXG 8
111
DB 1 YXG 3

RESULT 15
US-08-487-006-152
Sequence 152, Application US/08487006
Patent No. 5641861
GENERAL INFORMATION:

APPLICANT: Dooley, Colette T.
APPLICANT: Houghten, Richard A.
TITLE OF INVENTION: No. 5641861el Mu Opioid Receptor Ligands:
NUMBER OF SEQUENCES: 222
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,006
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-TP 1706
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 152:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
TOPOLOGY: linear
FEATURE:
NAME/KEY: Peptide
LOCATION: 2
OTHER INFORMATION: /note= "Xaa is (D)Nle."
FEATURE:
NAME/KEY: Peptide
LOCATION: 4
OTHER INFORMATION: /note= "Amino acid is amidated at
OTHER INFORMATION: the C-terminal."
US-08-487-006-152

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 YXG 8
111
DB 1 YXG 3

RESULT 16
US-08-487-006-153
Sequence 153, Application US/08487006
Patent No. 5641861
GENERAL INFORMATION:
APPLICANT: Dooley, Colette T.
APPLICANT: Houghten, Richard A.
TITLE OF INVENTION: No. 5641861el Mu Opioid Receptor Ligands:
NUMBER OF SEQUENCES: 222
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,006
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-TP 1706
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 153:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
TOPOLOGY: linear
FEATURE:
NAME/KEY: Peptide
LOCATION: 2
OTHER INFORMATION: /note="Xaa is (D)Nle."
FEATURE:
NAME/KEY: Peptide
LOCATION: 4
OTHER INFORMATION: /note="Amino acid is amidated at
OTHER INFORMATION: the C-terminal."
US-08-487-006-153

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 yxs 8
|||
Db 1 yxg 3

RESULT 17
US-08-454-950-7
Sequence 7, Application US/08454950
Patent No. 5690905
GENERAL INFORMATION:
APPLICANT: Zamora, Paul O.
APPLICANT: Rhodes, Buck A.
TITLE OF INVENTION: Peptide-Metal Ion
TITLE OF INVENTION: Pharmaceutical Preparation and Method
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rhomed Incorporated
STREET: 4261 Balloon Park
CITY: Albuquerque
STATE: NM
COUNTRY: U.S.A.
ZIP: 87109-5802
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB
MEDIUM TYPE: Storage
COMPUTER: IBM PC/XT/AT, IBM PS/2 or
COMPUTER: compatibles
OPERATING SYSTEM: PC-DOS or MS-DOS
SOFTWARE: Wordperfect 6.0a for Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/454,950
FILING DATE: 31-MAY-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/840,077
FILING DATE: 20-FEB-1992
APPLICATION NUMBER: 07/816,477
FILING DATE: 03-JAN-1992

ATTORNEY/AGENT INFORMATION:
NAME: Deborah A. Peacock
REGISTRATION NUMBER: 31,649
TELECOMMUNICATION INFORMATION:
TELEPHONE: (505) 242-9677
TELEFAX: (505) 243-2542
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-454-950-7

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wk 2
||
Db 2 wk 3

RESULT 18
US-08-434-761-3
Sequence 3, Application US/08434761
Patent No. 5698673
GENERAL INFORMATION:
APPLICANT: Blondelle, Sylvie E.
APPLICANT: Pinilla, Clemencia
APPLICANT: Elchler, Jutta
APPLICANT: Houghten, Richard A.
TITLE OF INVENTION: PEPTIDES HAVING ANTI-MICROBIAL,
TITLE OF INVENTION: HEMOLYTIC, ANTI-MELITTIN OR ANTI-ENZYMATIC ACTIVITY
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: CAMPBELL AND FLORES
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States of America
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/434,761
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/079,445
FILING DATE: 18-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-HP 9648
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
TOPOLOGY: linear
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1


```

OTHER INFORMATION: /label= Xaa
OTHER INFORMATION: /note= "Xaa-KfmcC"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /label= Xaa
OTHER INFORMATION: /note= "Xaa-any amino acid"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /note= "C-terminal amino acid is
OTHER INFORMATION: amldated"
US-08-434-761-3

Query Match      25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 wk 2
        11
        2 wk 3

Db

RESULT 19
US-08-338-890B-1
; Sequence 1, Application US/08338890B
; Patent No. 5700905
; GENERAL INFORMATION:
; APPLICANT: Hirschmann, Ralph
; APPLICANT: Spanevell, Rolando
; APPLICANT: Nutt, Ruth
; TITLE OF INVENTION: Synthetic Somatostatin Mimics
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz and No. 5700905rls
; STREET: One Liberty Place, 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: MS-DOS/MINDOWS
; SOFTWARE: WORDPERFECT SAVED AS ASCII (DOS) TEXT(*.*)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/338,890B
; FILING DATE: NOV. 14, 1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Malinoski, Lynn
; REGISTRATION NUMBER: 38,788
; REFERENCE/DOCKET NUMBER: UPN-2040
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215)568-3100
; TELEFAX: (215)568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-338-890B-1

Query Match      25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 wk 2
        11
        2 wk 3

Db
```

```

RESULT 20
US-07-923-260A-11
; Sequence 11, Application US/07923260A
; Patent No. 5719021
; GENERAL INFORMATION:
; APPLICANT: Inouye, Masayori
; TITLE OF INVENTION: PROTEIN ACTIVATION
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gerard J. Weiser
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/923,260A
; FILING DATE: 31-JUL-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 377.5638P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-07-923-260A-11

Query Match      25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 ay 6
        11
        2 ay 3

Db

RESULT 21
US-08-454-949-7
; Sequence 7, Application US/08454949
; Patent No. 5759516
; GENERAL INFORMATION:
; APPLICANT: Zamora, Paul O.
; APPLICANT: Rhodes, Buck A.
; TITLE OF INVENTION: Peptide-Metal Ion
; TITLE OF INVENTION: Pharmaceutical Preparation and Method
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhomed Incorporated
; STREET: 4261 Balloun Park
; CITY: Albuquerque
; STATE: NM
; COUNTRY: U.S.A.
; ZIP: 87109-5802
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB
; MEDIUM TYPE: Storage
; COMPUTER: IBM PC/XT/AT, IBM PS/2 or
; OPERATING SYSTEM: PC-DOS or MS-DOS
```

SOFTWARE: WordPerfect 6.0a for Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/454,949
FILING DATE: 31-MAY-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/840,077
FILING DATE: 20-FEB-1992
APPLICATION NUMBER: 07/816,477
FILING DATE: 03-JAN-1992
ATTORNEY/AGENT INFORMATION:
NAME: Deborah A. Peacock
REGISTRATION NUMBER: 31,649
TELECOMMUNICATION INFORMATION:
TELEPHONE: (505) 242-9677
TELEFAX: (505) 243-2542
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-454-949-7

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wk 2
11
Db 2 wk 3

RESULT 22
US-08-798-897-49
Sequence 49, Application US/08798897
Patent No. 5789201
GENERAL INFORMATION:
APPLICANT: Guastella, John
TITLE OF INVENTION: Genes Coding For Bcl-Y, a Bcl-2
TITLE OF INVENTION: Homologue
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/798,897
FILING DATE: February 11, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1483.0140001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid

STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-798-897-49

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ay 6
11
Db 3 AY 4

RESULT 23
US-08-463-224-2
Sequence 2, Application US/08463224
Patent No. 5807824
GENERAL INFORMATION:
APPLICANT: van Oostrum, Jan
APPLICANT: Boyar, William C.
APPLICANT: Galakatos, Nicholas G.
APPLICANT: Schmitz, Albert
APPLICANT: van Heeke, Gino
TITLE OF INVENTION: C5a Receptor Antagonists Having
SUBSTANTIALLY NO. 5807824AGONIST ACTIVITY
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lerner, David, Littenberg, Krumholz & Mentlik
STREET: 600 South Avenue West
CITY: Westfield
STATE: NJ
COUNTRY: USA
ZIP: 07090
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,224
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Foley, Shawn P.
TELECOMMUNICATION INFORMATION:
TELEPHONE: 908-654-5000
TELEFAX: 908-654-7866
TELEX: 139-125
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 4 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-463-224-2

Query Match 25.0%; Score 2; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ay 6
11
Db 3 AY 4

RESULT 24
US-08-463-377-2
Sequence 2, Application US/08463377
Patent No. 5837499
GENERAL INFORMATION:

APPLICANT: Van Oostrum, Jan
 APPLICANT: Boyar, William C.
 APPLICANT: Galakatos, Nicholas G.
 APPLICANT: Schmitz, Albert
 APPLICANT: van Hecke, Gino
 TITLE OF INVENTION: C5a Receptor Antagonists Having
 TITLE OF INVENTION: Substantially No. 5837499Agonist Activity
 NUMBER OF SEQUENCES: 67
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Lerner, David, Littenberg, Krumholz & Mentlik
 STREET: 600 South Avenue West
 CITY: Westfield
 STATE: NJ
 COUNTRY: USA
 ZIP: 07090
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/463,377
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Foley, Shawn P.
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 908-654-5000
 TELEFAX: 908-654-7866
 TELEX: 139-125
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-463-377-2

Query Match 25.0%; Score 2; DB 2; Length 4;
 Best Local Similarity 100.0%; Pred. No. 1.6e+05;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 5 ay 6
 Db 3 Ay 4

RESULT 25
 US-08-441-871-63
 Sequence 63, Application US/08441871
 Patent No. 5846765
 GENERAL INFORMATION:
 APPLICANT: Mathews, David J.
 APPLICANT: Zoller, Mark J.
 TITLE OF INVENTION: Identification of No. 5846765el Substrates
 NUMBER OF SEQUENCES: 152
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genentech, Inc.
 STREET: 460 Point San Bruno Blvd
 CITY: South San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94080
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patin (Genentech)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/441,871
 FILING DATE: 16-MAY-1995

CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/418928
 FILING DATE: 05-APR-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/161692
 FILING DATE: 03-DEC-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/864452
 FILING DATE: 06-APR-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/US91/09133
 FILING DATE: 03-DEC-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/743614
 FILING DATE: 09-AUG-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/715300
 FILING DATE: 14-JUN-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/683400
 FILING DATE: 10-APR-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/621667
 FILING DATE: 03-DEC-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: Winter, Daryl B.
 REGISTRATION NUMBER: 32,637
 REFERENCE/DOCKET NUMBER: 645P5C2D1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415/225-1249
 TELEFAX: 415/952-9881
 TELEX: 910/371-7168
 INFORMATION FOR SEQ ID NO: 63:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 US-08-441-871-63

Query Match 25.0%; Score 2; DB 2; Length 4;
 Best Local Similarity 100.0%; Pred. No. 1.6e+05;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 5 ay 6
 Db 2 Ay 3

Search completed: January 14, 2002, 07:57:30
 Job time: 375 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:58:36 ; Search time 45.4 Seconds
(without alignments)
13.423 Million cell updates/sec

Title: 09-185908-1c
Perfect score: 8
Sequence: 1 wxxxxfxg 8

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 50 summaries

Database : PIR_68:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	3	37.5	698	2 T13492	NADH dehydrogenase
2	2	25.0	5	2 PT0644	T-cell receptor be
3	2	25.0	7	2 S33244	neuromodulatory pe
4	2	25.0	7	2 S33246	neuromodulatory pe
5	2	25.0	7	2 PD0029	pey-kinin 1 - pena
6	2	25.0	8	2 A37521	R-phycocerythrin ga
7	2	25.0	8	2 S11078	glucose-6-phosphat
8	2	25.0	8	2 JS0316	leucokinin VI - Ma
9	2	25.0	8	2 A23967	leucoproteokin - M
10	2	25.0	8	2 D47393	neuropeptide calla
11	2	25.0	8	2 A42057	fibroblast growth
12	2	25.0	9	2 PT0315	Ig heavy chain CRD
13	2	25.0	10	1 RHLWGS	gonadoliberin - se
14	2	25.0	10	2 JCI367	thyroliberin poten
15	2	25.0	10	2 JN0024	neurokinin A - chl
16	2	25.0	10	2 D60787	sperm-activating p
17	2	25.0	10	2 S23307	neurokinin A - tai
18	2	25.0	10	2 S23186	neurokinin A - Acl
19	2	25.0	10	2 B61033	ranatachylkinin B -
20	2	25.0	10	2 C61033	neurokinin A - relat
21	2	25.0	10	2 S27178	neurokinin A - relat
22	2	25.0	10	2 I40032	type protein - Bac
23	2	25.0	10	2 S38305	lectin Gm2 alpha
24	2	25.0	10	2 D28027	protein P7 - curie
25	2	25.0	10	2 A43977	EMRFamlike-like pro
26	2	25.0	10	2 A49187	gonadotropin-rela
27	2	25.0	10	2 A37268	Ig heavy chain C r
28	2	25.0	10	2 PT0916	T-cell receptor be
29	2	25.0	10	2 T13838	cytochrome-c oxida

30	2	25.0	11	2 I60434	68kDa neurofilamen
31	2	25.0	12	1 U0G6M2	urotensin II - Ion
32	2	25.0	12	2 S42765	urotensin II - tel
33	2	25.0	12	2 JS0423	urotensin II-A pep
34	2	25.0	12	2 S26559	T-cell receptor be
35	2	25.0	12	2 S26554	T-cell receptor be
36	2	25.0	12	2 S69123	proton-translocat
37	2	25.0	12	2 B61497	seed protein ws-17
38	2	25.0	12	2 JS0424	urotensin II-B pep
39	2	25.0	12	2 JS10626	lipovitelalin - Afr
40	2	25.0	12	2 S74144	aggreca - bovine
41	2	25.0	12	2 PH1467	T-cell receptor be
42	2	25.0	12	2 PH1462	T-cell receptor be
43	2	25.0	12	2 PH1459	T-cell receptor be
44	2	25.0	12	2 PH1457	T-cell receptor be
45	2	25.0	12	2 PH0771	T-cell receptor be
46	2	25.0	12	2 PH1468	T-cell receptor be
47	2	25.0	13	2 PQ0445	urotensin II - lau
48	2	25.0	13	2 S08575	botulinum neurotox
49	2	25.0	13	2 A61514	glutathione trans
50	2	25.0	13	2 B56864	dipeptidyl-peptida

ALIGNMENTS

RESULT 1
T13492
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Nicodemia diversifolia chlorop
C/Species: chloroplast Nicodemia diversifolia
C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-Jan-2000
C/Accession: T13492
R:Olmstead, R.G.; Reeves, P.A.
Ann. Mo. Bot. Gard. 82, 176-193, 1995
A>Title: Evidence for the polyplyply of the Scrophulariaceae based on chloroplast rbcl
A/Reference number: 217559
A/Accession: T13492
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-698 <OLM>
A/Cross-references: EMBL:L36405; NID:g703196; PID:g703197; PIDN:AAA84496.1
C/Genetics:
A/Genome: chloroplast
A/Note: ndhf
C/Superfamily: NADH dehydrogenase (ubiquinone) chain 5
C/Keywords: chloroplast; membrane-associated complex; NAD; oxidoreductase

Query Match 37.5%; Score 3; DB 2; Length 698;
Best Local Similarity 100.0%; Pred. No. 3.2e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 sfxg 8
|||
Db 664 SFXG 667

RESULT 2
PT0644
T-cell receptor beta chain V-D-J region (111-1G) - mouse (fragment)
C/Species: Mus musculus (house mouse)
C/Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C/Accession: PT0644
R:Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A>Title: Junctional sequences of fetal T cell receptor beta chains have few N regions
A/Reference number: PT0509; MUID:91277601
A/Accession: PT0644
A/Status: translation not shown
A/Molecule type: mRNA
A/Residues: 1-5 <FEF>
A/Experimental source: newborn thymus, strain BALB/c
C/Keywords: T-cell receptor

Query Match 25.0%; Score 2; DB 2; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
11
Db 3 sf 4

RESULT 3
S33244
neuromodulatory peptide Wwamide-1 - giant African snail

C:Species: Achatina fulica (giant African snail)
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Jul-1997
C:Accession: S33244
R:Minakata, H.; Ikeda, T.; Muneoka, Y.; Kobayashi, M.; Nomoto, K.
FEBS Lett. 323, 104-108, 1993
A:Title: Wwamide-1, -2 and -3: novel neuromodulatory peptides isolated from ganglia of t
A:Reference number: S33244; MUID:93265912
A:Accession: S33244
A:Status: Preliminary
A:Molecule type: protein
A:Residues: 1-7 <MIN>

Query Match 25.0%; Score 2; DB 2; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
11
Db 1 wk 2

RESULT 4
S33246
neuromodulatory peptide Wwamide-3 - giant African snail

C:Species: Achatina fulica (giant African snail)
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Jul-1997
C:Accession: S33246
R:Minakata, H.; Ikeda, T.; Muneoka, Y.; Kobayashi, M.; Nomoto, K.
FEBS Lett. 323, 104-108, 1993
A:Title: Wwamide-1, -2 and -3: novel neuromodulatory peptides isolated from ganglia of t
A:Reference number: S33244; MUID:93265912
A:Accession: S33246
A:Status: Preliminary
A:Molecule type: protein
A:Residues: 1-7 <MIN>

Query Match 25.0%; Score 2; DB 2; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
11
Db 1 wk 2

RESULT 5
PD0029
pen-kinin 1 - penaeid shrimp (Penaeus vannamei) (fragment)

C:Species: Penaeus vannamei
C:Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 19-May-2000
C:Accession: PD0029
R:Nieto, J.; Veelaert, D.; Derua, R.; Waelkens, E.; Cersitlaens, A.; Coast, G.; Devreese, B.
Biochem. Biophys. Res. Commun. 248, 406-411, 1998
A:Title: Identification of one tachykinin- and two kinin-related peptides in the brain c
A:Reference number: PD0027; MUID:98342103
A:Accession: PD0029
A:Molecule type: protein

A:Residues: 1-7 <NIE>
C:Comment: This peptide belongs to myotropic neuropeptides.

Query Match 25.0%; Score 2; DB 2; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
11
Db 2 sf 3

RESULT 6
A37521
R-phycoerythrin gamma-E chain - red alga (Gastrocionium coulteri) (fragment)

C:Species: Gastrocionium coulteri
C:Date: 07-Mar-1988 #sequence_revision 07-Mar-1988 #text_change 23-Mar-1993
C:Accession: A37521; J22565
R:Klotz, A.V.; Glazer, A.N.
J. Biol. Chem. 260, 4856-4863, 1985
A:Title: Characterization of the bilin attachment sites in R-phycoerythrin.
A:Reference number: A22565; MUID:85182601
A:Accession: A37521
A:Molecule type: protein
A:Residues: 1-8 <KLO>

Query Match 25.0%; Score 2; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
11
Db 2 sf 3

RESULT 7
S11078
glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) - yeast (Pichia jadinii) (fragment)

C:Species: Pichia jadinii, Candida utilis
C:Date: 30-Jun-1991 #sequence_revision 30-Sep-1991 #text_change 05-Aug-1994
C:Accession: S11078
R:Eggestad, B.; Estolius, M.; Danielsson, O.; Persson, B.; Cederlund, E.; Kaiser, R.;
FEBS Lett. 269, 194-196, 1990
A:Title: Fast atom bombardment mass spectrometry and chemical analysis in determinati
A:Reference number: S11074; MUID:90355571
A:Accession: S11078
A:Molecule type: protein
A:Residues: 1-8 <EGE>
A>Note: the source is designated as Pichia jadinii
C:Keywords: acetylated amino end; oxidoreductase; pentose phosphate pathway
P:1/Modified site: acetylated amino end (Ser) #status experimental

Query Match 25.0%; Score 2; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
11
Db 4 sf 5

RESULT 8
J50316
leucokinin VI - Madelira cockroach

C:Species: Leucophaea maderae (Madelira cockroach)
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 20-Jun-2000
C:Accession: J50316
R:Holman, G.M.; Cook, B.J.; Nachman, R.J.
Comp. Biochem. Physiol. C 88, 27-30, 1987
A:Title: Isolation, primary structure, and synthesis of leucokinins V and VI: myotrop

A:Reference number: JS0315
A:Accession: JS0316
A:Molecule type: Protein
A:Residues: 1-8 <IOL>
C:Comment: Leucokinin, a family of cephalomyotropic peptides, stimulate contractile act
C:Keywords: amidated carboxyl end; cephalomyotropic peptide; pyroglutamic acid
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F:8/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 25.0%; Score 2; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
||
Db 3 sf 4

RESULT 9
A23967
Leucopyrokinin - *Medelira cockroach*
C:Species: *Leucophaea maderae* (*Medelira cockroach*)
C>Date: 31-Mar-1988 #sequence_revision 26-May-1994 #text_change 11-Jul-1997
C:Accession: A23967
R:Nachman, R.J.; Holman, G.M.; Cook, B.J.
B:Biochem. Biophys. Res. Commun. 137, 936-942, 1986
A:Title: Active fragments and analogs of the insect neuropeptide leucopyrokinin: structu
A:Reference number: A23967; MUID:86269041
A:Accession: A23967
A:Molecule type: Protein
A:Residues: 1-8 <NMC>
C:Keywords: amidated carboxyl end; neuropeptide; pyroglutamic acid
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F:8/Modified site: amidated carboxyl end (Leu) #status experimental

Query Match 25.0%; Score 2; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
||
Db 3 sf 4

RESULT 10
DA7393
neuropeptide calliostatin 4 - bluebottle fly (*Calliphora vomitoria*)
C:Species: *Calliphora vomitoria*
C>Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 28-Apr-1995
C:Accession: DA7393
R:Dave, H.; Johnson, A.H.; Scott, A.G.; Yu, C.G.; Yagci, K.J.; Tobe, S.S.; Thorpe, A.
Proc. Natl. Acad. Sci. U.S.A. 90, 2456-2460, 1993
A:Title: Calliostatins: neuropeptides from the blowfly *Calliphora vomitoria* with sequen
A:Reference number: A47393; MUID:93211980
A:Accession: DA7393
A:Status: preliminary
A:Molecule type: Protein
A:Residues: 1-8 <DUV>
A:Experimental source: thoracic ganglia
A:Note: sequence extracted from NCBI backbone (NCBIP:128479)

Query Match 25.0%; Score 2; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
||
Db 5 sf 6

RESULT 11
A42057
fibroblast growth factor receptor 1, secreted - mouse (fragment)
C:Species: *Mus musculus* (house mouse)
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 07-May-1999
C:Accession: A42057
R:Werner, S.; Duan, D.S.R.; de Vries, C.; Peters, K.G.; Johnson, D.E.; Williams, L.T.
Mol. Cell. Biol. 12, 82-88, 1992
A:Title: Differential splicing in the extracellular region of fibroblast growth facto
A:Reference number: A42057; MUID:92107200
A:Accession: A42057
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-8 <MER>
A:Cross-references: GB:M80363
C:Keywords: growth factor receptor

Query Match 25.0%; Score 2; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
||
Db 5 sf 6

RESULT 12
PT0315
Ig heavy chain CND3 region (clone 6-109) - human (fragment)
C:Species: *Homo sapiens* (man)
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Aug-1996
C:Accession: PT0315
R:Yamada, M.; Wasserman, R.; Reichard, B.A.; Shane, S.; Caton, A.J.; Rovera, G.
J. Exp. Med. 173, 395-407, 1991
A:Title: Preferential utilization of specific immunoglobulin heavy chain diversity an
A:Reference number: PT0222; MUID:91108337
A:Accession: PT0315
A:Molecule type: DNA
A:Residues: 1-9 <YAM>
A:Experimental source: B lymphocyte
C:Keywords: heterotetramer; immunoglobulin

Query Match 25.0%; Score 2; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
||
Db 7 sf 8

RESULT 13
RHLMGS
gonadoliberin - sea lamprey
N:Alternate names: gonadotropin releasing hormone (GnRH)
C:Species: *Petromyzon marinus* (sea lamprey)
C>Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 18-Mar-1997
C:Accession: A01412
R:Sherwood, N.M.; Sower, S.A.; Marshak, D.R.; Fraser, B.A.; Brownstein, M.J.
J. Biol. Chem. 261, 4812-4819, 1986
A:Title: Primary structure of gonadotropin-releasing hormone from lamprey brain.
A:Reference number: A01412; MUID:86168192
A:Accession: A01412
A:Molecule type: Protein
A:Residues: 1-10 <SHE>
C:Comment: This hormone was isolated from the brain.
C:Superfamily: gonadoliberin
C:Keywords: amidated carboxyl end; hormone; pyroglutamic acid
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F:10/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 25.0%; Score 2; DB 1; Length 10;
 Best Local Similarity 100.0%; Pred. No. 2.2e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
 ||
 Db 7 WK 8

RESULT 14
 JCI367

thyroliberin potentiating neuropeptide - bovine
 N:Alternate names: thyrotropin-releasing hormone potentiating peptide; TRH-potentiating
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 13-Mar-1997
 C:Accession: JCI367
 R:Bullet, M.; Ladrani, A.; Montagne, J.J.; Delfour, A.; Nicolas, P.
 Biochem. Biophys. Res. Commun. 189, 1110-1118, 1992
 A:Title: Isolation and amino acid sequence of the TRH-potentiating peptide from bovine
 A:Reference number: JCI367; MUID:93111999
 A:Accession: JCI367
 A:Molecule type: protein
 A:Residues: 1-10 <BOL>
 A:Experimental source: hypothalamus
 C:Comment: This neuropeptide corresponds to a region of the rat thyroliberin precursor
 C:Function:
 A:Description: potentiates thyroliberin-induced thyrotropin secretion
 C:Superfamily: thyroliberin precursor
 C:Keywords: hypothalamus; neuropeptide

Query Match 25.0%; Score 2; DB 2; Length 10;
 Best Local Similarity 100.0%; Pred. No. 2.2e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
 ||
 Db 1 SF 2

RESULT 15
 JN0024

neurokinin A - chicken
 C:Species: Gallus gallus (chicken)
 C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 18-Aug-2000
 C:Accession: JN0024
 R:Conlon, J.M.; Katsoulis, S.; Schmidt, W.E.; Thim, L.
 Regul. Pept. 20, 171-180, 1988
 A:Title: [Arg3]substance P and neurokinin A from chicken small intestine.
 A:Reference number: JN0023; MUID:88204263
 A:Accession: JN0024
 A:Molecule type: protein
 A:Residues: 1-10 <CON>
 C:Superfamily: substance P precursor
 C:Keywords: amidated carboxyl end; neuropeptide; tachykinin
 A:Reference number: JN0023; MUID:88204263
 F:10/Modified site: amidated carboxyl end (Met) #status predicted

Query Match 25.0%; Score 2; DB 2; Length 10;
 Best Local Similarity 100.0%; Pred. No. 2.2e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
 ||
 Db 5 SF 6

RESULT 16
 D60787

sperm-activating peptide (Ser-1, Ala-3, Gly-5 speract) - sea urchin (Hemicentrotus pulch
 C:Species: Hemicentrotus pulcherrimus
 C:Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 18-Aug-2000

C:Accession: D60787
 R:Suzuki, N.; Kajitani, H.; Nomura, K.; Garbers, D.L.; Yoshino, K.; Kurita, M.; Tanaka
 Comp. Biochem. Physiol. B 89, 687-693, 1988
 A:Title: Some more speract derivatives associated with eggs of sea urchins, pseudocen
 A:Reference number: A60787; MUID:88242184
 A:Accession: D60787
 A:Molecule type: protein
 A:Residues: 1-10 <SUZ>

C:Comment: This oligopeptide from egg jelly is one of several from this species, all
 at shows some, but not absolute, species restriction.
 C:Superfamily: unassigned animal peptides

Query Match 25.0%; Score 2; DB 2; Length 10;
 Best Local Similarity 100.0%; Pred. No. 2.2e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
 ||
 Db 1 SF 2

RESULT 17
 S23307

neurokinin A - rainbow trout
 C:Species: Oncorhynchus mykiss (rainbow trout)
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 18-Aug-2000
 C:Accession: S23307
 R:Jensen, J.; Conlon, J.M.
 Eur. J. Biochem. 206, 659-664, 1992
 A:Title: Substance-P-related and neurokinin-A-related peptides from the brain of the
 A:Reference number: S23186; MUID:92298992
 A:Accession: S23307
 A>Status: preliminary
 A:Molecule type: protein
 A:Residues: 1-10 <JEN>
 C:Superfamily: unassigned animal peptides

Query Match 25.0%; Score 2; DB 2; Length 10;
 Best Local Similarity 100.0%; Pred. No. 2.2e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
 ||
 Db 5 SF 6

RESULT 18
 S23186

neurokinin A - Atlantic cod
 C:Species: Gadus morhua (Atlantic cod)
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 18-Aug-2000
 C:Accession: S23186
 R:Jensen, J.; Conlon, J.M.
 Eur. J. Biochem. 206, 659-664, 1992
 A:Title: Substance-P-related and neurokinin-A-related peptides from the brain of the
 A:Reference number: S23186; MUID:92298992
 A:Accession: S23186
 A:Molecule type: protein
 A:Residues: 1-10 <JEN>
 A:Experimental source: brain
 C:Function:
 A:Description: may play a physiological role in the regulation of cardiovascular and
 A>Note: neurokinin A is derived by post-translational processing of preprotachykinin
 C:Superfamily: unassigned animal peptides
 C:Keywords: neuropeptide; amidated carboxyl end; tachykinin
 F:10/Modified site: amidated carboxyl end (Met) #status predicted

Query Match 25.0%; Score 2; DB 2; Length 10;
 Best Local Similarity 100.0%; Pred. No. 2.2e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
||
Db 5 SF 6

RESULT 19

B61033
Rana tachykinin B - bullfrog
C:Species: Rana catesbeiana (bullfrog)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 18-Aug-2000
C:Accession: B61033; J60427
R:Kangawa, K.; Kozawa, H.; Hino, J.; Minamino, N.; Matsuo, H.
Regul. Pept. 42(Suppl.1), S12, 1992
A:Title: Isolation of four novel tachykinins from frog (Rana catesbeiana) brain and intestine
A:Reference number: A61033
A:Accession: B61033
A:Molecule type: protein
A:Residues: 1-10 <KAN>
R:Kozawa, H.; Hino, J.; Minamino, N.; Kangawa, K.; Matsuo, H.
Biochem. Biophys. Res. Commun. 177, 588-595, 1991
A:Title: Isolation of four novel tachykinins from frog (Rana catesbeiana) brain and intestine
A:Reference number: J60426; MUID:91254337
A:Accession: J60427

A:Molecule type: protein
A:Residues: 1-10 <KOZ>
C:Superfamily: unassigned animal peptides
C:Keywords: amidated carboxyl end; neuropeptide
F:10/Modified site: amidated carboxyl end (Met) #status experimental

Query Match 25.0%; Score 2; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 2.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
||
Db 5 SF 6

RESULT 20

C61033
Rana tachykinin C - bullfrog
C:Species: Rana catesbeiana (bullfrog)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 18-Aug-2000
C:Accession: C61033; J60428
R:Kangawa, K.; Kozawa, H.; Hino, J.; Minamino, N.; Matsuo, H.
Regul. Pept. 42(Suppl.1), S12, 1992
A:Title: Isolation of four novel tachykinins from frog (Rana catesbeiana) brain and intestine
A:Reference number: A61033
A:Accession: C61033
A:Molecule type: protein
A:Residues: 1-10 <KAN>
R:Kozawa, H.; Hino, J.; Minamino, N.; Kangawa, K.; Matsuo, H.
Biochem. Biophys. Res. Commun. 177, 588-595, 1991
A:Title: Isolation of four novel tachykinins from frog (Rana catesbeiana) brain and intestine
A:Reference number: J60426; MUID:91254337
A:Accession: J60428

A:Molecule type: protein
A:Residues: 1-10 <KOZ>
C:Superfamily: unassigned animal peptides
C:Keywords: amidated carboxyl end; neuropeptide
F:10/Modified site: amidated carboxyl end (Met) #status experimental

Query Match 25.0%; Score 2; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 2.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
||
Db 5 SF 6

RESULT 21

S27178
neurokinin A-related peptide - laughing frog
C:Species: Rana ridibunda (laughing frog)
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 18-Aug-2000
C:Accession: S27178
R:Wang, Y.; Badgery-Parker, T.; Lovas, S.; Chartrel, N.; Vaudry, H.; Burcher, E.; Con
Biochem. J. 287, 827-832, 1992
A:Title: Primary structure and receptor-binding properties of a neurokinin A-related peptide
A:Reference number: S27178; MUID:93075037
A:Accession: S27178
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-10 <MAN>
C:Superfamily: unassigned animal peptides

Query Match 25.0%; Score 2; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 2.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
||
Db 5 SF 6

RESULT 22

I40032
type protein - Bacillus amyloliquefaciens (fragment)
C:Species: Bacillus amyloliquefaciens
C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 15-Oct-1999
C:Accession: I40032
R:Yoshimura, K.; Uemura, J.; Seki, T.; Oshima, Y.
J. Bacteriol. 159, 905-912, 1984
A:Title: Construction of a promoter-probe vector for Bacillus subtilis host by using a reference number: I40032; MUID:9506754
A:Accession: I40032
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-10 <RES>
A:Cross-references: GB:K02661; NID:g143775; PIDN:AB05353.1; PID:g143776

Query Match 25.0%; Score 2; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 2.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
||
Db 5 SF 6

RESULT 23

S38305
lectin GNL2 alpha chain - kidney bean (fragment)
C:Species: Phaseolus vulgaris (kidney bean)
C:Date: 19-May-1994 #sequence_revision 27-Feb-1997 #text_change 13-Mar-1997
C:Accession: S38305
R:Kamemura, K.; Furutachi, Y.; Umekawa, H.; Takahashi, T.
Biochim. Biophys. Acta 1158, 181-188, 1993
A:Title: Purification and characterization of novel lectins from Great Northern bean, Phaseolus communis
A:Reference number: S38304; MUID:94002183
A:Accession: S38305
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-10 <KAN>

Query Match 25.0%; Score 2; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 2.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
11
Db 5 SF 6

RESULT 24

D28027
protein p7 - curled-leaved tobacco (fragment)
C:Species: Nicotiana glauca (curled-leaved tobacco)
C:Date: 19-May-1989 #sequence_revision 19-May-1989 #text_change 18-Jun-1993
C:Accession: D28027
R:Baum, G.; De Loose, M.; Inze, D.; Van Montagu, M.; Vandekerckhove, J.
Proc. Natl. Acad. Sci. U.S.A. 84, 4806-4810, 1987
A:Title: Alterations in the phenotype of plant cells studied by NH2-terminal amino acid-
A:Reference number: A94167
A:Accession: D28027
A:Molecule type: protein
A:Residues: 1-10 <BAU>

Query Match 25.0%; Score 2; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 2.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
11
Db 5 SF 6

RESULT 25

A43977
FMRFamide-like protein - tobacco hornworm
C:Species: Manduca sexta (tobacco hornworm)
C:Date: 11-Feb-1993 #sequence_revision 11-Feb-1993 #text_change 17-Mar-1999
C:Accession: A43977
R:Kingan, T.G.; Teplow, D.B.; Phillips, J.M.; Riehm, J.P.; Rao, K.R.; Hildebrand, J.G.;
Peptides 11, 849-856, 1990
A:Title: A new peptide in the FMRFamide family isolated from the CNS of the hawkmoth, Ma
A:Reference number: A43977; MUID:91045350
A:Accession: A43977
A:Molecule type: protein
A:Residues: 1-10 <KIN>
C:Keywords: amidated carboxyl end; neuropeptide; pyroglutamic acid
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F:10/Modified site: amidated carboxyl end (Phe) #status experimental

Query Match 25.0%; Score 2; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 2.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
11
Db 6 SF 7

Search completed: January 14, 2002, 07:58:37
Job time: 367 sec

```

RESULT 23
TKNB_RANRI STANDARD: PRT: 10 AA.
AC P29135:
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE NEUROKININ A.
OS Rana ridibunda (Laughing frog) (Marsh frog).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
NC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Ranidae; Rana.
NX NCBI_TaxID=8406;
RN [1]
RP SEQUENCE.
RC TISSUE=Skin;
RX MEDLINE=93075037; PubMed=1332683;
RA Wang Y., Badgerly-Parker T., Lomas S., Chitrel N., Vaudry H.,
RT Burcher E., Conlon J.M.;
RT "Primary structure and receptor-binding properties of a neurokinin A-
related peptide from frog gut.";
RL Biochem. J. 287:827-832(1992).
CC -1- FUNCTION: TACHYKININS ARE ACTIVE PEPTIDES WHICH EXCITE NEURONS,
SECRETAGOGUES, AND CONTRACT (DIRECTLY OR INDIRECTLY) MANY SMOOTH
MUSCLES.
CC -1- SIMILARITY: BELONGS TO THE TACHYKININ FAMILY.
DR PIR: S27178; S27178.
DR InterPro: IPR002040; Tachykinin.
DR PROSITE: PS00267; TACHYKININ, 1.
KW Tachykinin; Neuropeptide; Amidation; Amphibian skin.
FT MOD_RES 10 10 AMIDATION.
SQ SEQUENCE 10 AA; 1160 MW; 526B407059D5BAA7 CRC64;

Query Match
Best Local Similarity 100.0%; Score 2; DB 1; Length 10;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
DB 5 sf 6

RESULT 24
TKNC_RANCA STANDARD: PRT: 10 AA.
AC P22690;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE RANAVTACHYKININ C (RTK C).
OS Rana catesbeiana (Bull Frog).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
NC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Ranidae; Rana.
NX NCBI_TaxID=8400;
RN [1]
RP SEQUENCE, AND SYNTHESIS.
RC TISSUE=Intestine;
RX MEDLINE=91254337; PubMed=2043143;
RA Kozawa H., Hino J., Minamino N., Kangawa K., Matsuo H.;
RT "Isolation of four novel tachykinins from frog (Rana catesbeiana)
brain and intestine.";
RL Biochem. Biophys. Res. Commun. 177:588-595(1991).
RN [2]
RP SEQUENCE.
RC TISSUE=Intestine;
RX MEDLINE=94023216; PubMed=8210506;
RA Kangawa K., Kozawa H., Hino J., Minamino N., Matsuo H.;
RT "Four novel tachykinins in frog (Rana catesbeiana) brain and
intestine.";
RL Regl. Pept. 46:81-88(1993).
CC -1- FUNCTION: TACHYKININS ARE ACTIVE PEPTIDES WHICH EXCITE NEURONS,
EVOLVE BEHAVIORAL RESPONSES, ARE POTENT VASODILATORS AND

```

```

CC SECRETAGOGUES, AND CONTRACT (DIRECTLY OR INDIRECTLY) MANY SMOOTH
MUSCLES.
CC -1- SIMILARITY: BELONGS TO THE TACHYKININ FAMILY.
DR PIR: JE0428; JE0428.
DR PIR: C61033; C61033.
DR InterPro: IPR002040; Tachykinin.
DR PROSITE: PS00267; TACHYKININ, 1.
KW Tachykinin; Neuropeptide; Amidation.
FT MOD_RES 10 10 AMIDATION.
SQ SEQUENCE 10 AA; 1086 MW; 3A3A407059D5BDC7 CRC64;

Query Match
Best Local Similarity 100.0%; Score 2; DB 1; Length 10;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
DB 5 sf 6

RESULT 25
MHBI_KLEPN STANDARD: PRT: 11 AA.
AC P80580;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE MALEYLPYRUVATE ISOMERASE (EC 5.2.1.4) (FRAGMENT).
GN MHBI.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
NC Klebsiella.
NX NCBI_TaxID=573;
RN [1]
RP SEQUENCE.
RX MEDLINE=96349117; PubMed=8760924;
RA Robson N.D., Parrott S., Cooper R.A.;
RT "In vitro formation of a catabolic plasmid carrying Klebsiella
pneumoniae DNA that allows growth of Escherichia coli K-12 on 3-
hydroxybenzoate.";
RL Microbiology 142:2115-2120(1996).
CC -1- CATALYTIC ACTIVITY: 3-MALEYLPYRUVATE = 3-FUMARYLPYRUVATE.
KW Isomerase.
FT NON_TER 11 11
SQ SEQUENCE 11 AA; 1387 MW; 1EE0E2DD49C9D5AB CRC64;

Query Match
Best Local Similarity 100.0%; Score 2; DB 1; Length 11;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
DB 5 sf 6

```

Search completed: January 14, 2002, 08:08:22
 Job time: 707 sec

Query Match 25.0%; Score 2; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sf 6
DB 1
1 sf 2

RESULT 20
TKNB_CHICK STANDARD; PRT; 10 AA.
P19851;
01-FEB-1991 (Rel. 17, Created)
01-FEB-1991 (Rel. 17, Last sequence update)
30-MAY-2000 (Rel. 39, Last annotation update)
NEUROKININ A (SUBSTANCE K) (NEUROMEDIN 1).
Gallus gallus (Chicken).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
Gallus.
NCBI_TaxID=9031;
[1]
SEQUENCE.
Tissue=Intestine; PubMed=2452461;
MEDLINE=88204263; Conlon J.M., Katsoulis S., Schmidt W.E., Thim L.,
[Arg3]substance P and neurokinin A from chicken small intestine.";
Regul. Pept. 20:171-180(1988).
-I- FUNCTION: TACHYKININS ARE ACTIVE PEPTIDES WHICH EXCITE NEURONS,
EVOKE BEHAVIORAL RESPONSES, ARE POTENT VASODILATORS AND
SECRETAGOGUES, AND CONTRACT (DIRECTLY OR INDIRECTLY) MANY SMOOTH
MUSCLES.
-I- SIMILARITY: BELONGS TO THE TACHYKININ FAMILY.
PIR: JN0024; JN0024.
InterPro: IPR002040; Tachykinin.
DR PROSITE: PS00267; TACHYKININ; 1.
KW Tachykinin; Neuropeptide; Amidation.
FT MOD_RES 10
RT SEQUENCE 10 AA; 1134 MW; 8A6B4062C9D5B41 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sf 6
DB 5 sf 6

RESULT 21
TKNB_ONCMY STANDARD; PRT; 10 AA.
P28500;
01-DEC-1992 (Rel. 24, Created)
01-DEC-1992 (Rel. 24, Last sequence update)
30-MAY-2000 (Rel. 39, Last annotation update)
NEUROKININ A (SUBSTANCE K) (NEUROMEDIN 1).
Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri), and
Gadus morhua (Atlantic cod).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
NCBI_TaxID=8022, 8049;
[1]
SEQUENCE.
Tissue=Brain;
MEDLINE=92298992; PubMed=1376687;
RX Jensen J., Conlon J.M.,
RA "Substance-P-related and neurokinin-A-related peptides from the brain
of the cod and trout.";
Eur. J. Biochem. 206:659-664(1992).

CC -I- FUNCTION: TACHYKININS ARE ACTIVE PEPTIDES WHICH EXCITE NEURONS,
EVOKE BEHAVIORAL RESPONSES, ARE POTENT VASODILATORS AND
SECRETAGOGUES, AND CONTRACT (DIRECTLY OR INDIRECTLY) MANY SMOOTH
MUSCLES.
CC -I- SIMILARITY: BELONGS TO THE TACHYKININ FAMILY.
PIR: S23186; S23186.
DR InterPro: IPR002040; Tachykinin.
DR PROSITE: PS00267; TACHYKININ; 1.
KW Tachykinin; Neuropeptide; Amidation.
FT MOD_RES 10
RT SEQUENCE 10 AA; 1145 MW; 136B4062C9D5B440 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sf 6
DB 5 sf 6

RESULT 22
TKNB_RANCA STANDARD; PRT; 10 AA.
P22689;
01-AUG-1991 (Rel. 19, Created)
01-AUG-1991 (Rel. 19, Last sequence update)
30-MAY-2000 (Rel. 39, Last annotation update)
RANMATACHYKININ B (RTK B).
Rana catesbeiana (Bull Frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Ranidae; Rana.
NCBI_TaxID=8400;
[1]
SEQUENCE, AND SYNTHESIS.
RP Tissue=Brain;
RC MEDLINE=91254337; PubMed=2043143;
RX Kozawa H., Hino J., Minamino N., Kangawa K., Matsuo H.,
RT "Isolation of four novel tachykinins from frog (Rana catesbeiana)
brain and intestine.";
RL Biochem. Biophys. Res. Commun. 177:588-595(1991).
RN [2]
RP SEQUENCE.
RC Tissue=Intestine;
RX MEDLINE=94023216; PubMed=8210506;
RA Kangawa K., Kozawa H., Hino J., Minamino N., Matsuo H.,
RT "Four novel tachykinins in frog (Rana catesbeiana) brain and
intestine.";
RL Regul. Pept. 46:81-88(1993).
CC -I- FUNCTION: TACHYKININS ARE ACTIVE PEPTIDES WHICH EXCITE NEURONS,
EVOKE BEHAVIORAL RESPONSES, ARE POTENT VASODILATORS AND
SECRETAGOGUES, AND CONTRACT (DIRECTLY OR INDIRECTLY) MANY SMOOTH
MUSCLES.
CC -I- SIMILARITY: BELONGS TO THE TACHYKININ FAMILY.
PIR: JE0427; JE0427.
DR PIR: B61033; B61033.
DR InterPro: IPR002040; Tachykinin.
DR PROSITE: PS00267; TACHYKININ; 1.
KW Tachykinin; Neuropeptide; Amidation.
FT MOD_RES 10
RT SEQUENCE 10 AA; 1210 MW; 917E556B59D5B4B5 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sf 6
DB 5 sf 6

CC FLIGHT BEHAVIOR PATTERNS.
CC -1- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE) FAMILY.
CC PIR: AA3977; AA3977.
DR Amidation: Neuropeptide.
KW MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 10 10 AMIDATION.
SQ SEQUENCE 10 AA; 1247 MW; D3CA5229D5B1F2D2 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
11
DB 6 sf 7

RESULT 17
ID GONL_PETMA STANDARD; PRT; 10 AA.
AC P04378;
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE GONADOLIBERIN I (GONADOTROPIN-RELEASING HORMONE I) (GNRH-I) (LUTIBERIN I).
DE Petromyzon marinus (Sea Lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
OC NCBI_TaxID=7757;
RN (1)
RP SEQUENCE.
RC TISSUE: Brain;
RX MEDLINE=86168192; PubMed=3514603;
RA Sherwood N.M., Sower S.A., Marshak D.R., Fraser B.A., Brownstein M.J.;
RT "Primary structure of gonadotropin-releasing hormone from lamprey brain."
RL J. Biol. Chem. 261:4812-4819(1986).
CC -1- FUNCTION: STIMULATES THE SECRETION OF BOTH LUTEINIZING AND FOLLICLE-STIMULATING HORMONES.
CC -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.
CC PIR: A01412; RHLMS.
DR InterPro: IPR002012; GNRH.
DR Pfam: PF00446; GNRH.1.
DR PROSITE: PS00473; GNRH.1.
KW Hormone; Amidation; Hypothalamus.
FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 10 10 AMIDATION.
SQ SEQUENCE 10 AA; 1244 MW; 1E4B36237B1735AB CRC64;

Query Match 25.0%; Score 2; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
11
DB 7 wk 8

RESULT 18
ID GON3_PETMA STANDARD; PRT; 10 AA.
AC P30948;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GONADOLIBERIN III (GONADOTROPIN-RELEASING HORMONE III) (GNRH-III) (LULIBERIN III).
DE Petromyzon marinus (Sea Lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;

OC Petromyzontiformes; Petromyzontidae; Petromyzon.
OX NCBI_TaxID=7757;
RN (1)
RP SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=93178316; PubMed=8440174;
RA Sower S.A., Chiang Y.-C., Lovas S., Conlon J.M.;
RT "Primary structure and biological activity of a third gonadotropin-releasing hormone from lamprey brain."
RL Endocrinology 132:1125-1131(1993).
CC -1- FUNCTION: STIMULATES THE SECRETION OF BOTH LUTEINIZING AND FOLLICLE-STIMULATING HORMONES.
CC -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.
DR InterPro: IPR002012; GNRH.
DR Pfam: PF00446; GNRH.1.
DR PROSITE: PS00473; GNRH.1.
KW Hormone; Amidation; Hypothalamus.
FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 10 10 AMIDATION.
SQ SEQUENCE 10 AA; 1277 MW; 284B36237A1F5A3 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
11
DB 7 wk 8

RESULT 19
ID RCA_PINPS STANDARD; PRT; 10 AA.
AC P81084;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE RIBULOSE BISPHOSPHATE CARBOXYLASE/OXYGENASE (RUBISCO ACTIVASE) (RA) (WATER STRESS RESPONSIVE PROTEIN 4) (FRAGMENT).
DE Pinus pinaster (Maritime pine).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=71647;
RN (1)
RP SEQUENCE.
RC TISSUE=Needle;
RX MEDLINE=98418576; PubMed=9747804;
RA Costa P., Bahrmann N., Frigerio J.-M., Kremer A., Plomion C.;
RT "Water-deficit-responsive proteins in maritime pine.";
RL Plant Mol. Biol. 38:587-596(1998).
RN (2)
RP SEQUENCE.
RC TISSUE=Needle;
RX MEDLINE=99274088; PubMed=10344291;
RA Costa P., Plomion C., Bauw G., Dubos C., Bahrmann N., Kremer A., Frigerio J.-M., Plomion C.;
RT "Separation and characterization of needle and xylem maritime pine proteins.";
RL Electrophoresis 20:1098-1108(1999).
CC -1- FUNCTION: ACTIVATION OF RUBISCO (RIBULOSE-1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE; EC 4.1.1.39) INVOLVES THE ATP-DEPENDENT CARBOXYLATION OF THE EPSILON-AMINO GROUP OF LYSINE LEADING TO A CARBAMATE STRUCTURE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST STROMA (BY SIMILARITY).
CC -1- INDUCTION: BY WATER-STRESS.
CC -1- SIMILARITY: BELONGS TO THE RUBISCO ACTIVASE FAMILY.
KW Chloroplast; ATP-binding.
FT NON_TER 1 1
FT NON_TER 10 10
SQ SEQUENCE 10 AA; 1171 MW; C0A506D2C72B1E46 CRC64;

CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
 CC Blaberidae; Blaberidae; Leucophaea.
 CC NCBI_TaxID=6988;
 CC (1)
 CC SEQUENCE.
 RP MEDLINE=86269041; PubMed=3015140;
 RA Nachman R.J., Holman G.M., Cook B.J.;
 RT "Active fragments and analogs of the insect neuropeptide
 RL leucopyrokinin: structure-function studies."
 RL Biochem. Biophys. Res. Commun. 137:936-942(1986).
 CC (12)
 CC SEQUENCE, AND SYNTHESIS.
 RP TISSUE=Head;
 RC MEDLINE=87052651; PubMed=2877794;
 RX Holman G.M., Cook B.J., Nachman R.J.;
 RT "Primary structure and synthesis of a blocked myotropic
 RT neuropeptide isolated from the cockroach, *Leucophaea maderae*."
 RL Comp. Biochem. Physiol. 85C:219-224(1986).
 CC (1)
 CC -1- FUNCTION: MEDIATES VISCERAL MUSCLE CONTRACTILE ACTIVITY
 CC (MYOTROPIC ACTIVITY).
 CC -1- MISCELLANEOUS: AN ANALOG WITHOUT THE N-TERMINAL PCA RESIDUE WAS
 CC SYNTHESIZED AND FOUND TO EXHIBIT GREATER ACTIVITY (144%) THAN THE
 CC PARENT NEUROPEPTIDE. THE PORTION OF THE SEQUENCE OF LPK MOST
 CC CRITICAL FOR THE MYOTROPIC PROPERTIES IS LIMITED TO THE
 CC PENTAPEPTIDE FRAGMENT FTPLRL.
 CC -1- SIMILARITY: BELONGS TO THE PYROKININ FAMILY.
 DR PIR: A23967; A23967.
 DR Interpro: IPR001484; Pyrokinin.
 DR PROSITE: PS00539; PYROKININ; 1.
 CC NEUROPEPTIDE: Amidation; Pyrokinin.
 FT MOD_RES 1 1 PYROLIDONE CARBOXYLIC ACID.
 FT MOD_RES 8 8 AMIDATION.
 FT SEQUENCE 8 AA: 949 MW: 92341771A9D5A1B6 CRC64:

Query Match 25.0%; Score 2; DB 1; Length 8;
 Best Local Similarity 100.0%; Pred. No. 1e+05;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sf 6
 II
 Db 3 SF 4

RESULT 14
 FARP_CALSI STANDARD; PRT; 9 AA.
 AC P38495;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE FMRFAMIDE-LIKE NEUROPEPTIDE.
 OS Callinectes sapidus (Blue crab).
 CC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
 CC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
 CC Eubrachyura; Portunoidae; Portunidae; Callinectes.
 CC NCBI_TaxID=6763;
 CC (1)
 CC SEQUENCE.
 RP MEDLINE=92270479; PubMed=1815216;
 RA Krajinak K.G.;
 RT "The identification and structure-activity relations of a
 RT cardioactive FMRFamide-related peptide from the blue crab *Callinectes*
 RT *sapidus*." 12:1295-1302(1991).
 RL Peptides 12:1295-1302(1991).
 CC -1- FUNCTION: CARDIOACTIVE PEPTIDE.
 CC -1- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
 CC FAMILY.
 CC NEUROPEPTIDE: Amidation.
 FT MOD_RES 9 9 AMIDATION.
 FT SEQUENCE 9 AA: 1159 MW: 134F0729D5A4045B CRC64:

Query Match 25.0%; Score 2; DB 1; Length 9;
 Best Local Similarity 100.0%; Pred. No. 1e+05;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sf 6
 II
 Db 5 SF 6

RESULT 15
 AL19_CARMA STANDARD; PRT; 10 AA.
 AC P81822;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE CARCINOSTATIN 19.
 OS Carcinus maenas (Common shore crab) (Green crab).
 CC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
 CC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
 CC Eubrachyura; Portunoidae; Portunidae; Carcinus.
 CC NCBI_TaxID=6759;
 CC (1)
 CC SEQUENCE.
 RP TISSUE=Cerebral ganglion, and Thoracic ganglion;
 RC MEDLINE=98121193; PubMed=9461295;
 RA Duye H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
 RA Thorpe A.;
 RT "Isolation and identification of multiple neuropeptides of the
 RT allatostatin superfamily in the shore crab *Carcinus maenas*."
 RL Eur. J. Biochem. 250:727-734(1997).
 CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
 CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
 CC NEUROPEPTIDE: Amidation; Multigene family.
 FT MOD_RES 10 10 AMIDATION (POTENTIAL).
 FT SEQUENCE 10 AA: 1101 MW: 96687CD5AB569AB1 CRC64:

Query Match 25.0%; Score 2; DB 1; Length 10;
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sf 6
 II
 Db 7 SF 8

RESULT 16
 FARP_MANSE STANDARD; PRT; 10 AA.
 AC P18523;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE FMRFAMIDE-LIKE NEUROPEPTIDE.
 OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
 CC Spingioidea; Spingidae; Spinginae; Manduca.
 CC NCBI_TaxID=7150;
 CC (1)
 CC SEQUENCE.
 RP MEDLINE=91045350; PubMed=2235684;
 RX Kingan T.G., Teplov D.B., Phillips J.M., Riehm J.P., Rao K.R.,
 RA Hildebrand J.G., Homberg U., Kammer A.E., Jardine I., Griffin P.R.,
 RA Hunt D.F.;
 RT "A new peptide in the FMRFamide family isolated from the CNS of the
 RT hawkmoth, *Manduca sexta*." 11:849-856(1990).
 RL Peptides 11:849-856(1990).
 CC -1- FUNCTION: INCREASES THE FORCE OF NEURALLY EVOKED CONTRACTIONS IN
 CC THE MAJOR POWER-PRODUCING FLIGHT MUSCLES, THE DORSAL LONGITUDINAL
 CC MUSCLES AND SO IS LIKELY TO PLAY A ROLE IN SUSTAINING OR PROMOTING

CC BEHAVE AS POTENT INHIBITORS OF PERISTALTIC MOVEMENTS. MAY ALSO
 CC FULFILL A NEUROHORMONAL ROLE ON MUSCLES OF THE GUT AND HEART.
 CC -1- TISSUE SPECIFICITY: BRAIN, SUBESOPHAGEAL GANGLION, RETROCEREBRAL
 CC COMPLEX, THORACICO-ABDOMINAL GANGLION, PERIPHERAL NEUROSECRETORY
 CC SYSTEM AND INTESTINE.
 CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
 DR PIR: D47393; D47393.
 KW Neuropeptide; Amidation.
 FT MOD_RES 8
 FT UNSURE 1 OR N.
 SQ SEQUENCE 8 AA: 954 MW: D32879D5AB47740A CRC64;

Query Match 25.0%; Score 2; DB 1; Length 8;
 Best Local Similarity 100.0%; Pred. No. 1e+05;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
 II
 DB 5 sf 6

RESULT 10
 ALL4_CYPDO
 ID ALL4_CYPDO STANDARD; PRT; 8 AA.
 AC P82155;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE CYDIASTATIN 4.
 OS Cydia pomonella (Coddling moth).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
 CC Tortricidae; Tortricidae; Olethreutinae; Cydia.
 OX NCBI_TaxID=82600;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Larva;
 RX MEDLINE=98054539; PubMed=9392829;
 RA Duvé H., Johnsen A.H., Maestro J.-L., Scott A.G., Winstanley D.,
 RA Davey M., East P.D., Thorpe A.;
 RA "Lepidopteran peptides of the allatostatin superfamily.";
 RL Peptides 18:1301-1309(1997).
 CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
 KW Neuropeptide; Amidation.
 FT MOD_RES 8
 FT UNSURE 1
 SQ SEQUENCE 8 AA: 910 MW: 922879D5AB47740D CRC64;

Query Match 25.0%; Score 2; DB 1; Length 8;
 Best Local Similarity 100.0%; Pred. No. 1e+05;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
 II
 DB 5 sf 6

RESULT 11
 LCK4_LEUMA
 ID LCK4_LEUMA STANDARD; PRT; 8 AA.
 AC P21143;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 01-MAY-1991 (Rel. 18, Last annotation update)
 DE LEUCOKININ IV (L-IV).
 OS Leucophaea maderae (Madelira cockroach).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
 CC Blaberoidea; Blaberoidea; Leucophaea.
 OX NCBI_TaxID=6988;
 RN [1]
 RP AND SYNTHESIS.

RC TISSUE=Head;
 RA Holman G.M., Cook B.J., Nachman R.J.;
 RT "Primary structure and synthesis of two additional neuropeptides
 RT from Leucophaea maderae: members of a new family of
 RT Cephalomyotriptides."
 RL Comp. Biochem. Physiol. 84C:271-276(1986).
 CC -1- FUNCTION: THIS CEPHALOMYOTROPIC PEPTIDE STIMULATES CONTRACTILE
 CC ACTIVITY OF COCKROACH PROTODEUM (HINDGUT).
 CC -1- SIMILARITY: TO THE OTHER LEUCOKININS.
 KW Neuropeptide; Amidation.
 FT MOD_RES 8
 FT UNSURE 1
 SQ SEQUENCE 8 AA: 906 MW: DC6365B1E9D5BDDA CRC64;

Query Match 25.0%; Score 2; DB 1; Length 8;
 Best Local Similarity 100.0%; Pred. No. 1e+05;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
 II
 DB 3 sf 4

RESULT 12
 LCK6_LEUMA
 ID LCK6_LEUMA STANDARD; PRT; 8 AA.
 AC P19988;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE LEUCOKININ VI (L-VI).
 OS Leucophaea maderae (Madelira cockroach).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
 CC Blaberoidea; Blaberoidea; Leucophaea.
 OX NCBI_TaxID=6988;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Head;
 RX MEDLINE=87052651; PubMed=2877794;
 RA Holman G.M., Cook B.J., Nachman R.J.;
 RT "Isolation, primary structure, and synthesis of leucokinin V and VI:
 RT myotropic peptides of Leucophaea maderae.";
 RL Comp. Biochem. Physiol. 88C:27-30(1987).
 CC -1- FUNCTION: THIS CEPHALOMYOTROPIC PEPTIDE STIMULATES CONTRACTILE
 CC ACTIVITY OF COCKROACH PROTODEUM (HINDGUT).
 CC -1- SIMILARITY: TO THE OTHER LEUCOKININS, AND TO MANDUCA SEXTA AND
 CC HELIOTHIS ZEA ADIPOKINETIC HORMONE.
 DR PIR: JS0316; JS0316.
 KW Neuropeptide; Amidation.
 FT MOD_RES 1
 FT UNSURE 1
 SQ SEQUENCE 8 AA: 935 MW: 9D6365B1E9D5A5A6 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 8;
 Best Local Similarity 100.0%; Pred. No. 1e+05;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
 II
 DB 3 sf 4

RESULT 13
 LPK_LEUMA
 ID LPK_LEUMA STANDARD; PRT; 8 AA.
 AC P13049;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE LEUCOPYROKININ (LPK) (LEM-PK).
 OS Leucophaea maderae (Madelira cockroach).

```

RESULT 6
AL17_CARMA STANDARD: PRT: 8 AA.
AC P81820;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CARCINOSTATIN 17.
DE Carcinus maenas (Common shore crab) (Green crab).
CC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
CC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
CC Eubrachyura; Portunoidae; Portunidae; Carcinus.
CC NCBI_TaxID=6759;
RN [1]
RP TISSUE=Cerebral ganglion, and Thoracic ganglion;
RC MEDLINE=98121193; PubMed=9461295;
RA Duvé H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA Thorpe A.;
RT "Isolation and identification of multiple neuropeptides of the
RT allatostatin superfamily in the shore crab Carcinus maenas.";
RT Eur. J. Biochem. 250:727-734(1997).
CC -1- FUNCTION: MAY ACT AS A NEUROTANSMITTER OR NEUROMODULATOR.
CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
CC MOD_RES 8 AA: 919 MW; C82879D5AB569AB5 CRC64;
SQ SEQUENCE 8 AA: 919 MW; C82879D5AB569AB5 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sf 6
Db 5 SF 6

RESULT 7
AL18_CARMA STANDARD: PRT: 8 AA.
AC P81821;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CARCINOSTATIN 18.
DE Carcinus maenas (Common shore crab) (Green crab).
CC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
CC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
CC Eubrachyura; Portunoidae; Portunidae; Carcinus.
CC NCBI_TaxID=6759;
RN [1]
RP TISSUE=Cerebral ganglion, and Thoracic ganglion;
RC MEDLINE=98121193; PubMed=9461295;
RA Duvé H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RA Thorpe A.;
RT "Isolation and identification of multiple neuropeptides of the
RT allatostatin superfamily in the shore crab Carcinus maenas.";
RT Eur. J. Biochem. 250:727-734(1997).
CC -1- FUNCTION: MAY ACT AS A NEUROTANSMITTER OR NEUROMODULATOR.
CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
CC MOD_RES 8 AA: 919 MW; C82879D5AB569AB5 CRC64;
SQ SEQUENCE 8 AA: 919 MW; C82879D5AB569AB5 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sf 6
Db 5 SF 6

```

```

QY 5 sf 6
Db 5 SF 6

RESULT 8
ALL3_CYPDPO STANDARD: PRT: 8 AA.
ID ALL3_CYPDPO
AC P82154;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CYDIASTATIN 3.
DE Cydia pomonella (Codling moth).
CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
CC Eukaryota; Metazoa; Arthropoda; Lepidoptera; Glossata; Ditrysia;
CC Tortricoidae; Tortricidae; Olethreutinae; Cydia.
CC NCBI_TaxID=82600;
RN [1]
RP TISSUE=Larva;
RC MEDLINE=98054539; PubMed=9392829;
RA Duvé H., Johnsen A.H., Maestro J.-L., Scott A.G., Winstanley D.,
RA Davey M., East P.D., Thorpe A.;
RT "Peptidoprotein peptides of the allatostatin superfamily.";
RL Peptides 18:1301-1309(1997).
CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
CC MOD_RES 8 AA: 926 MW; C82879D5AB477415 CRC64;
SQ SEQUENCE 8 AA: 926 MW; C82879D5AB477415 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sf 6
Db 5 SF 6

```

```

RESULT 9
ALL4_CALVO STANDARD: PRT: 8 AA.
ID ALL4_CALVO
AC P41840;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CALLATOSTATIN 4 (Leu-CALLATOSTATIN 4).
DE Calliphora vomitoria (Blue blowfly).
CC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
CC Eukaryota; Metazoa; Arthropoda; Diptera; Brachycera; Muscomorpha;
CC Oestroidae; Calliphoridae; Calliphora.
CC NCBI_TaxID=27454;
RN [1]
RP TISSUE=Thoracic ganglion;
RC MEDLINE=93211980; PubMed=8460157;
RA Duvé H., Johnsen A.H., Scott A.G., Yu C.G., Yagi K.J., Tobe S.S.,
RA Thorpe A.;
RT "Callatostatins: neuropeptides from the blowfly Calliphora vomitoria
RT with sequence homology to cockroach allatostatins.";
RT Proc. Natl. Acad. Sci. U.S.A. 90:2456-2460(1993).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=94291167; PubMed=8020069;
RX Duvé H., Thorpe A.;
RA "Distribution and functional significance of Leu-callatostatins in
RA the blowfly Calliphora vomitoria.";
RL Cell Tissue Res. 276:367-379(1994).
CC -1- FUNCTION: MAY ACT AS A NEUROTANSMITTER OR NEUROMODULATOR AND PLAY
CC A ROLE IN THE INTEGRATION OF INFORMATION WITHIN THE BRAIN. MAY BE
CC INVOLVED IN THE CONTROL OF VISCERAL MUSCLES DUE TO ITS ABILITY TO

```


11] SEQUENCE FROM N.A.
RX MEDLINE:94357160; PubMed-8076592;
RA Parra-Lopez C., Lin R., Aspedon A., Groisman E.A.;
RT "A Salmonella protein that is required for resistance to
antimicrobial peptides and transport of potassium.";
RL EMBO J. 13:3964-3972(1994).
CC -1- FUNCTION: CHANNEL THAT OPENS IN RESPONSE TO STRETCH FORCES IN THE
MEMBRANE LIPID BILAYER. MAY PARTICIPATE IN THE REGULATION OF
OSMOTIC PRESSURE CHANGES WITHIN THE CELL (BY SIMILARITY).
CC -1- SUBUNIT: HOMOPENTAMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
(BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE MSCL FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X80501; CAA5664.1; -.
DR StyGene: SGI0531; mscL.
DR InterPro: IPR001185; MSCL.
DR PROSITE: PS01327; MSCL; PARTIAL.
KW Transmembrane; Inner membrane; Ionic channel.
FT NON_TER
SQ SEQUENCE 7 AA: 901 MW: 69CB133059D5B6F0 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 7;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
DB 2 sf 3

RESULT 3
ID WMA2_ACHFU STANDARD; PRT; 7 AA.
AC P35920;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE WMAWIDE-2.
OS Achatina fulica (Giant African snail).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;
OC Achatinacea; Achatinidae; Achatina.
OX NCBI_TaxID=6530;
RN [1]
RP SEQUENCE.
RC TISSUE:Ganglion;
RX MEDLINE:93265912; PubMed-8495720;
RA Minakata H., Ikeda T., Muneoka Y., Kobayashi M., Nomoto K.;
RT "WMAWIDE-1, -2 and -3: novel neuromodulatory peptides isolated from
ganglia of the African giant snail, Achatina fulica.";
RL FEBS Lett. 323:104-108(1993).
KW Neuropeptide; Amidation.
FT MOD_RES
SQ SEQUENCE 7 AA: 964 MW: 7362D5B686D32310 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 7;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
DB 1 wk 2

RESULT 4
ID WMA3_ACHFU STANDARD; PRT; 7 AA.
AC P35921;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE WMAWIDE-3
OS Achatina fulica (Giant African snail).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;
OC Achatinacea; Achatinidae; Achatina.
OX NCBI_TaxID=6530;
RN [1]
RP SEQUENCE.
RC TISSUE:Ganglion;
RX MEDLINE:93265912; PubMed-8495720;
RA Minakata H., Ikeda T., Muneoka Y., Kobayashi M., Nomoto K.;
RT "WMAWIDE-1, -2 and -3: novel neuromodulatory peptides isolated from
ganglia of the African giant snail, Achatina fulica.";
RL FEBS Lett. 323:104-108(1993).
DR PIR: S33244; S33244.
KW Neuropeptide; Amidation.
FT MOD_RES
SQ SEQUENCE 7 AA: 965 MW: 7362D5B69B132310 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 7;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wk 2
DB 1 wk 2

RESULT 5
ID AL15_CARMA STANDARD; PRT; 8 AA.
AC P81818;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CARCINUSSTRATIN 15.
OS Carcinus maenas (Common shore crab) (Green crab).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC Eudrachyura; Portunoidae; Portunidae; Carcinus.
OX NCBI_TaxID=6759;
RN [1]
RP SEQUENCE.
RC TISSUE=Cerebral ganglion, and Thoracic ganglion;
RX MEDLINE:98121193; PubMed-9461295;
RA Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
Thorpe A.;
RT "Isolation and identification of multiple neuropeptides of the
allatostatin superfamily in the shore crab Carcinus maenas.";
RL Eur. J. Biochem. 250:727-734(1997).
CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW Neuropeptide; Amidation; Multigene family.
FT MOD_RES
SQ SEQUENCE 8 AA: 811 MW: 922879D5AB47687D CRC64;

Query Match 25.0%; Score 2; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
DB 5 sf 6

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

09-185908-1c

Run on: January 14, 2002, 08:08:21 ; Search time 30.66 Seconds
(without alignments)
9.567 Million cell updates/sec

Title: 09-185908-1c
Perfect score: 8
Sequence: 1 wxxsfxg 8

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size: 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

Database: SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2	25.0	5	1 ALI4_CARMA	P81817 carcinius ma
2	2	25.0	7	1 MSCL_SALTY	P39446 salmomella
3	2	25.0	7	1 WMA2_ACHFU	P35920 achatina fu
4	2	25.0	7	1 WMA3_ACHFU	P35921 achatina fu
5	2	25.0	8	1 ALI5_CARMA	P81818 carcinius ma
6	2	25.0	8	1 ALI7_CARMA	P81820 carcinius ma
7	2	25.0	8	1 ALI8_CARMA	P81821 carcinius ma
8	2	25.0	8	1 ALI3_CYDPO	P82154 cydia pomon
9	2	25.0	8	1 ALI4_CALVO	P81840 calliphora
10	2	25.0	8	1 ALI4_CYDPO	P82155 cydia pomon
11	2	25.0	8	1 LCK4_LEUMA	P21143 leucophaea
12	2	25.0	8	1 LCK6_LEUMA	P19988 leucophaea
13	2	25.0	8	1 LPK_LEUMA	P13049 leucophaea
14	2	25.0	9	1 FARP_CALSI	P38495 callinectes
15	2	25.0	10	1 ALI9_CARMA	P81822 carcinius ma
16	2	25.0	10	1 FARP_MANSE	P18523 manduca sex
17	2	25.0	10	1 GON1_PETMA	P04378 petromyzon
18	2	25.0	10	1 GON3_PETMA	P30948 petromyzon
19	2	25.0	10	1 RCA_PINPS	P81084 pinus pinas
20	2	25.0	10	1 TKNB_CHICK	P19851 gallus gall
21	2	25.0	10	1 TKNB_ONCMY	P28500 oncorhynch
22	2	25.0	10	1 TKNB_RANCA	P22689 rana catesb
23	2	25.0	10	1 TKNB_RANCA	P22689 rana catesb
24	2	25.0	10	1 TKNB_RANCA	P22689 rana catesb
25	2	25.0	11	1 MHR1_KLEPN	P80580 klebsiella
26	2	25.0	12	1 UR2B_CATCO	P04559 catostomus
27	2	25.0	12	1 UR2B_CATCO	P04559 catostomus
28	2	25.0	12	1 UR2B_CATCO	P04559 catostomus
29	2	25.0	12	1 UR2_GILMI	P01147 gilllichthys
30	2	25.0	12	1 UR2_POLSP	P81022 polyodon sp
31	2	25.0	12	1 UR2_SCYCA	P32490 scyllorhynch
32	2	25.0	12	1 V25K_MSSV	P83004 white spot
33	2	25.0	14	1 LECB_PSOOC	P22584 psophocarpu

34	2	25.0	14	1 MAST_PARID	PA2716 parapolypia
35	2	25.0	14	1 MAST_POLJA	P01517 polistes ja
36	2	25.0	14	1 MAST_VESXA	P01515 vespa xanth
37	2	25.0	14	1 SMS1_MYOSC	P20750 myoxocephal
38	2	25.0	14	1 SMS_ALIMI	P13185 alligator m
39	2	25.0	15	1 ASP1_LACSN	P82648 lactobacilli
40	2	25.0	15	1 ATP2_PINPS	P81663 pinus pinas
41	2	25.0	15	1 CBP2_PROAT	P19628 protopetrus
42	2	25.0	15	1 DCM1_PSECA	P19920 pseudomonas
43	2	25.0	15	1 FIBA_ANAPL	P12801 anas platyr
44	2	25.0	15	1 HSL1_PINPS	P81083 pinus pinas
45	2	25.0	15	1 LEC1_PSOOC	P22582 psophocarpu
46	2	25.0	15	1 LEC2_PSOOC	P22583 psophocarpu
47	2	25.0	15	1 LEC3_PSOOC	P22583 psophocarpu
48	2	25.0	15	1 MALT_BACTO	P80072 bacillus th
49	2	25.0	15	1 MILT_ONCKE	P81037 oncorhynch
50	2	25.0	15	1 UC25_MAIZE	P80631 zea mays (m

ALIGNMENTS

RESULT 1
ALI4_CARMA STANDARD; PRT; 5 AA.
ID ALI4_CARMA
AC P81817;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CARCINUSTATIN 14, (Common shore crab) (Green crab).
OS Carcinus maenas (Common shore crab) (Green crab).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC Eubrachyura; Portunoidae; Portunidae; Carcinus.
OX NCBI_TaxID=6759;
RN [1]
RP SEQUENCE.
RC TISSUE=Cerebral ganglion, and Thoracic ganglion;
RX MEDLINE=98121193; PubMed=9461295;
RA Drove H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
Thorp A.;
RT "Isolation and identification of multiple neuropeptides of the
RT allatostatin superfamily in the shore crab Carcinus maenas.";
RL Eur. J. Biochem. 250:727-734 (1997).
CC -1- PUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW Neuropeptide; Amidation; Multigene family.
FT MOD_RES 5
FT AMIDATION (POTENTIAL).
SQ SEQUENCE 5 AA; 586 MW; 672879DSAB300000 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 5;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sf 6
Db 11
2 sf 3

RESULT 2
MSCL_SALTY STANDARD; PRT; 7 AA.
ID MSCL_SALTY
AC P39446;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE LARGE-CONDUCTANCE MECHANOSENSITIVE CHANNEL (FRAGMENT).
GN MSCL.
OS Salmomella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmomella.
OX NCBI_TaxID=602;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 08:07:31 ; Search time 81.98 Seconds
(without alignments)
14,274 Million cell updates/sec

Title: 09-185908-1C
Perfect score: 8
Sequence: 1 wkxxsfxxg 8

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

- Database : SPTREMBL_17.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phase.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3	37.5	23	6 Q9TRC6	Q9trc6 canis famil
2	3	37.5	56	2 Q51622	Q51622 escherichia
3	3	37.5	144	8 Q9G8X4	Q9g8x4 toxostoma c
4	3	37.5	153	8 Q9T5K3	Q9t5k3 haemonchus
5	3	37.5	271	5 Q9N7E2	Q9n7p2 leishmania
6	3	37.5	290	8 Q9Z2G6	Q9zzg6 phyllotis d
7	3	37.5	297	8 Q9TDH9	Q9tdh9 agelaius ru
8	3	37.5	347	8 Q634I8	Q634i8 habia rubic
9	3	37.5	353	5 Q9NNI3	Q9nni3 leishmania
10	3	37.5	379	8 Q9TF74	Q9tf74 spermophilu
11	3	37.5	385	5 Q9GWR8	Q9gwr8 leishmania
12	3	37.5	413	5 Q44100	Q44100 drosophila
13	3	37.5	475	8 Q9TID6	Q9tid6 pieris flor
14	3	37.5	510	8 Q9BAX6	Q9bax6 gongora gra
15	3	37.5	515	8 Q9BAX0	Q9bax0 gongora sph
16	3	37.5	608	8 Q326E5	Q326e5 nicotinia d
17	3	37.5	707	8 Q9MVF7	Q9mvf7 muntingia c
18	3	37.5	717	8 Q9MVG6	Q9mvg6 luehea seem
19	2	25.0	7	2 Q07354	Q07354 synechococc

20	2	25.0	7	8 Q95945	Q95945 saccharomyc
21	2	25.0	8	3 P87225	P87225 saccharomyc
22	2	25.0	8	4 Q9HCQ0	Q9hcq0 homo sapien
23	2	25.0	8	5 P82685	P82685 periplaneta
24	2	25.0	8	5 P82686	P82686 periplaneta
25	2	25.0	8	5 P82687	P82687 periplaneta
26	2	25.0	8	11 Q62721	Q62721 rattus norv
27	2	25.0	9	2 P72345	P72345 pseudomonas
28	2	25.0	9	4 P78484	P78484 homo sapien
29	2	25.0	9	4 Q9UQW0	Q9uq0 homo sapien
30	2	25.0	9	4 Q9UQW0	Q9uq0 homo sapien
31	2	25.0	9	4 Q9BQT4	Q9bqt4 homo sapien
32	2	25.0	9	11 P97889	P97889 rattus norv
33	2	25.0	9	12 Q67605	Q67605 squash leaf
34	2	25.0	9	12 Q67606	Q67606 squash leaf
35	2	25.0	10	2 Q44693	Q44693 bacillus am
36	2	25.0	10	2 Q9L5W6	Q9l5w6 liberibacte
37	2	25.0	10	6 Q9TRC1	Q9trc1 bos taurus
38	2	25.0	10	8 P92576	P92576 bipes bipor
39	2	25.0	10	8 Q9XMB4	Q9xmb4 aegilops ta
40	2	25.0	10	12 Q66190	Q66190 avian infec
41	2	25.0	10	12 Q84140	Q84140 influenza a
42	2	25.0	11	2 Q56972	Q56972 yersinia pe
43	2	25.0	11	8 Q9G649	Q9g649 otocryptis
44	2	25.0	11	8 Q9G646	Q9g646 sitana pont
45	2	25.0	11	8 Q9G643	Q9g643 calotes cal
46	2	25.0	11	8 Q9G640	Q9g640 calotes cey
47	2	25.0	11	8 Q9G634	Q9g634 calotes llo
48	2	25.0	11	8 Q9G631	Q9g631 calotes nig
49	2	25.0	11	8 Q9G628	Q9g628 calotes mys
50	2	25.0	11	8 Q9G625	Q9g625 calotes ver

ALIGNMENTS

RESULT 1					
Q9TRC6					
ID	Q9TRC6	PRELIMINARY;	PRT;	23 AA.	
AC	Q9TRC6;				
DT	01-MAY-2000 (TRENBLrel. 13, Created)				
DT	01-MAY-2000 (TRENBLrel. 13, Last sequence update)				
DT	01-JUN-2000 (TRENBLrel. 14, Last annotation update)				
DE	CYTCHROME P450 DPB-1 ISOFORM (FRAGMENT).				
OS	Canis familiaris (Dog)				
OC	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.				
OX	NCBI_TaxID=9615;				
RN	[1]				
RP	SEQUENCE.				
RX	MEDLINE=94198820; PubMed=8185738;				
RA	Shiraga T., Iwasaki K., Nozaki K., Tamura T., Yamazoe Y., Kato R.,				
RA	Takanaka A.;				
RT	"Isolation and characterization of four cytochrome P450 isozymes from				
RT	untreated and phenobarbital-treated beagle dogs.";				
RL	Biol. Pharm. Bull. 17:22-28(1994).				
SQ	SEQUENCE 23 AA; 2638 MW; AF868888FEF8029F8 CRC64;				
Query Match	37.5%;	Score 3;	DB 6;	Length 23;	
Best Local Similarity	100.0%;	Pred. No. 5.8e+02;			
Matches	4;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	5 sfxg 8				
Db	6 SFXG 9				
RESULT 2					
Q51622					
ID	Q51622	PRELIMINARY;	PRT;	56 AA.	
AC	Q51622;				
DT	01-NOV-1996 (TRENBLrel. 01, Created)				

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE MOB9 ORF.
OS Escherichia coli.
OG Plasmid Cole1.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85261261; PubMed=2991225;
RA Chan P.T., Ohmori H., Tomizawa J., LeBowitz J.;
RT "Nucleotide sequence and gene organization of Cole1 DNA.";
RL J. Biol. Chem. 260:8925-8935(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89210827; PubMed=3149585;
RA Stirling C.J., Szatmari G., Stewart G., Smith M.C., Sherratt D.J.;
RT "The arginine repressor is essential for plasmid-stabilizing site-specific recombination at the Cole1 cer locus.";
RL EMBO J. 7:4389-4395(1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=91123196; PubMed=1846858;
RA Inoue N., Uchida H.;
RT "Transcription and initiation of Cole1 DNA replication in Escherichia coli K-12.";
RL J. Bacteriol. 173:1208-1214(1991).
DR EMBL; J01566; AAB59139.1; -.
KW Plasmid.
SQ SEQUENCE 56 AA; 6254 MW; 35ABA84389978BEF8 CRC64;

Query Match 37.5%; Score 3; DB 2; Length 56;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 sfxg 8
||||
Db 26 SFXG 29

RESULT 3
Q9G8X4
ID Q9G8X4 PRELIMINARY; PRT; 144 AA.
AC Q9G8X4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CYTOCHROME B (FRAGMENT).
GN CYTB.
OS Toxostoma curvirostre.
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Passeriformes; Sturnidae; Toxostoma.
OX NCBI_TaxID=99878;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBTH52NM;
RA Zink R.M., Blackwell-Rago R.C.;
RT "Species limits and recent population history of the Curve-billed Thrasher.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY BOUND TO THE PROTEIN (BY SIMILARITY).
CC -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B, CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME B/B6 FAMILY.
DR EMBL; AF287539; AAB31529.1; -.

DR InterPro; IPR000179; Cyt_b_b6.
DR Pfam; PF00033; cytochrome_b_n; 1.
DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
KW Electron transport; Heme; Mitochondrion; Respiratory chain;
KW Transmembrane.
FT NON_TER 1
FT NON_TER 144
SQ SEQUENCE 144 AA; 15972 MW; 4B0320E8A892724D CRC64;

Query Match 37.5%; Score 3; DB 8; Length 144;
Best Local Similarity 100.0%; Pred. No. 2.3e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 sfxg 8
||||
Db 107 SFXG 110

RESULT 4
Q9T5K3
ID Q9T5K3 PRELIMINARY; PRT; 153 AA.
AC Q9T5K3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT 4 (FRAGMENT).
GN ND4.
OS Haemochus placei.
OG Mitochondrion.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
OC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.
OX NCBI_TaxID=6290;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99083438; PubMed=9866206;
RA Blouin M.S., Yowell C.A., Courtney C.H., Dame J.B.;
RT "Substitution bias, rapid saturation, and the use of mtDNA for nematode systematics.";
RL Mol. Biol. Evol. 15:1719-1727(1998).
CC -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC -!- SIMILARITY: TO NADH-UBIQUINONE/PLASTOQUINONE (COMPLEX I), VARIOUS CHAINS.

DR EMBL; AF070801; AAC98219.1; -.
DR InterPro; IPR001750; Oxidored_q1.
DR Pfam; PF00361; oxidored_q1; 1.
KW Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
FT NON_TER 1
SQ SEQUENCE 153 AA; 18040 MW; 726825467487DB8F CRC64;

Query Match 37.5%; Score 3; DB 8; Length 153;
Best Local Similarity 100.0%; Pred. No. 2.4e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 sfxg 8
||||
Db 119 SFXG 122

RESULT 5
Q9N7P2
ID Q9N7P2 PRELIMINARY; PRT; 271 AA.
AC Q9N7P2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE POSSIBLE PROBABLE ECF-FAMILY SIGMA FACTOR (FRAGMENT).
GN LM28.157.
OS Leishmania major.
OC Eukaryota; Eudlenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]

```
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL390935; CAC00895.1;
FT NON_TER 1 271
FT NON_TER 271 271
SQ SEQUENCE 271 AA; 30356 MW; B59E63D2D321BFF4 CRC64;

Query Match 37.5%; Score 3; DB 5; Length 271;
Best Local Similarity 100.0%; Pred. No. 3.7e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sfxg 8
DB 154 SFXG 157

RESULT 6
Q92ZG6 PRELIMINARY; PRT; 290 AA.
AC Q92ZG6;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CYTOCHROME B (FRAGMENT).
GN CYTB.
OS Phyllotis darwini (Darwin's leaf-eared mouse).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Sigmodontinae;
OC Phyllotis.
OX NCBI_TaxID=56232;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UNM NK27558;
RA Stepan S.J.;
RT "Phylogenetic relationships and species limits within Phyllotis
RT (Rodentia: Sigmodontinae): concordance between mtDNA sequence and
RT morphology.";
RL J. Mammal. 79:0-0(1998).
CC -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN (BY SIMILARITY).
CC -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME B/B6 FAMILY.
DR EMBL: U86819; AAD1238.1;
DR InterPro: IPR000179; Cyt_b_b6.
DR Pfam: PF00033; cytochrome_b_n; 1.
DR Pfam: PF00033; cytochrome_b_n; 1.
DR PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
DR PROSITE: PS00193; CYTOCHROME_B_QO; UNKNOWN_1.
DR Electron transport; Heme; Mitochondrion; Respiratory chain;
KW Transmembrane.
FT NON_TER 1 290
FT NON_TER 290 290
SQ SEQUENCE 290 AA; 32572 MW; BCEAF9D0DBED2771 CRC64;

Query Match 37.5%; Score 3; DB 8; Length 290;
Best Local Similarity 100.0%; Pred. No. 3.9e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sfxg 8
DB 136 SFXG 139
```

```
RESULT 7
Q9TDH9 PRELIMINARY; PRT; 297 AA.
AC Q9TDH9;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CYTOCHROME B (FRAGMENT).
GN CYTB.
OS Agelaius ruficapillus.
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Passeriformes; Icteridae; Agelaius.
OX NCBI_TaxID=84775;
RN [1]
RP SEQUENCE FROM N.A.
RA Lanyon S.M., Omland K.E.;
RT "A molecular phylogeny of the blackbirds (Icteridae): five lineages
RT revealed by cytochrome-b sequence data.";
RL Auk 116:629-639(1999).
CC -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN (BY SIMILARITY).
CC -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME B/B6 FAMILY.
DR EMBL: AF089009; AAF02229.1;
DR InterPro: IPR000179; Cyt_b_b6.
DR Pfam: PF00032; cytochrome_b_c; 1.
DR Pfam: PF00033; cytochrome_b_n; 1.
DR PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
DR PROSITE: PS00193; CYTOCHROME_B_QO; UNKNOWN_1.
DR Electron transport; Heme; Mitochondrion; Respiratory chain;
KW Transmembrane.
FT NON_TER 1 297
FT NON_TER 297 297
SQ SEQUENCE 297 AA; 33128 MW; 8B62C22B82EE2ED5 CRC64;

Query Match 37.5%; Score 3; DB 8; Length 297;
Best Local Similarity 100.0%; Pred. No. 4e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sfxg 8
DB 100 SFXG 103

RESULT 8
O63418 PRELIMINARY; PRT; 347 AA.
AC O63418;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CYTOCHROME B (FRAGMENT).
GN CYTB.
OS Habia rubica.
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Passeriformes; Fringillidae;
OC Emberizinae; Habia.
OX NCBI_TaxID=62201;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98086775; PubMed=9417892;
RA Burns K.J.;
RT "Molecular systematics of tanagers (Thraupinae): evolution and
RT biogeography of a diverse radiation of neotropical birds.";
RL Mol. Phylogenet. Evol. 8:334-348(1997).
```

CC -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: QH(2) + 2 FERRICYTOCHROME C = O + 2
CC FERROCYTOCHROME C.
CC -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN (BY SIMILARITY).
CC -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
CC EMBL; AF006233; AAC05054.1; -.
DR InterPro: IPR001600; Galanin.
DR Pfam: PF00032; cytochrome_b_c1.
DR PRINTS; PR00273; GALANIN.
DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR PROSITE; PS00193; CYTOCHROME_B_QO; UNKNOWN_1.
KW Electron transport; Heme; Mitochondrion; Respiratory chain;
KW Transmembrane. 1
FT NON_TER 1
SQ SEQUENCE 347 AA; 38616 MW; 81208E4E94D37A8E CRC64;

Query Match 37.5%; Score 3; DB 8; Length 347;
Best Local Similarity 100.0%; Pred. No. 4.5e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 sfxg 8
Db 107 SFXG 110

RESULT 9
ID Q9NNI3 PRELIMINARY; PRT; 353 AA.
AC Q9NNI3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE POSSIBLE 75 KDA INVARIANT SURFACE GLYCOPROTEIN (FRAGMENT).
GN LM15.181.
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FRIEDLIN;
RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL160371; CAC00227.1; -.
FT NON_TER 1
FT NON_TER 353
SQ SEQUENCE 353 AA; 38247 MW; 3E427E2171CC7F24 CRC64;

Query Match 37.5%; Score 3; DB 5; Length 353;
Best Local Similarity 100.0%; Pred. No. 4.6e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 sfxg 8
Db 258 SFXG 261

RESULT 10
ID Q9TF74 PRELIMINARY; PRT; 379 AA.
AC Q9TF74;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 17, Last annotation update)

DE CYTOCHROME B.
GN CYTB.
OS Spemophilus erythrogenys (red-cheeked ground squirrel).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutelestomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Scuridae; Scurinae;
OC Spemophilus.
OX NCBI_TaxID=99840;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S136;
RA Harrison R.G., Sherman P.W., Yensen E., Hoffmann R.S.,
RA Bogdanowicz S.M.;
RT "A molecular phylogeny of ground squirrels and prairie dogs."
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN (BY SIMILARITY).
CC -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
CC EMBL; AF157875; AAD50159.1; -.
DR InterPro: IPR000179; Cyt_b_b6.
DR Pfam: PF00032; cytochrome_b_c1.
DR Pfam: PF00033; cytochrome_b_n; 1.
DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
DR PROSITE; PS00193; CYTOCHROME_B_QO; UNKNOWN_1.
KW Electron transport; Heme; Mitochondrion; Respiratory chain;
KW Transmembrane.
SQ SEQUENCE 379 AA; 42795 MW; B0C35BAFE3118854 CRC64;

Query Match 37.5%; Score 3; DB 8; Length 379;
Best Local Similarity 100.0%; Pred. No. 4.8e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 sfxg 8
Db 139 SFXG 142

RESULT 11
ID Q9GWR8 PRELIMINARY; PRT; 385 AA.
AC Q9GWR8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 43.0 KDA PROTEIN (FRAGMENT).
GN LM12.753.
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FRIEDLIN;
RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,
RA Oliver K.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL390114; CAC02470.1; -.
DR InterPro: IPR001522; Desaturase.
DR PRINTS; PR00075; FACDSDSATRASE.
DR ProDom; PD002221; Desaturase; 1.
KW Hypothetical protein.
FT NON_TER 1
FT NON_TER 385
SQ SEQUENCE 385 AA; 43000 MW; A79EF6194CFA960A CRC64;

Query Match 37.5%; Score 3; DB 5; Length 385;


```

Best Local Similarity 100.0%; Pred. No. 4.9e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sfxg 8
DB 247 SFXG 250

RESULT 12
O44100
ID O44100 PRELIMINARY; PRT; 413 AA.
AC O44100
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ENOLASE (EC 4.2.1.11) (2-PHOSPHOGLYCERATE DEHYDRATASE) (2-PHOSPHO-D-
GLYCERATE HYDRO-LYASE) (FRAGMENT).
GN ENO.
OS Drosophila pseudoobscura (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7237;
RN [1]
RP SEQUENCE FROM N.A.
RA Zeng L.-W., Comeron J.M., Chen B., Kreitman M.;
RL Genetica 0:0-0(1997).
CC -1- CATALYTIC ACTIVITY: 2-PHOSPHO-D-GLYCERATE = PHOSPHOENOLPYRUVATE +
H(2)O.
CC -1- COFACTOR: MAGNESIUM IS REQUIRED FOR CATALYSIS AND FOR STABILIZING
THE DIMER (BY SIMILARITY).
CC -1- PATHWAY: GLYCOLYSIS.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: TO THE ENOLASE FAMILY.
DR EMBL; AF025805; AAB87890.1; -.
DR HSP; P56232; IPDZ.
DR FlyBase; FBgn0023296; Dpse\Eno.
DR InterPro; IPR000941; Enolase.
DR Pfam; PF01113; enolase; 1.
DR PRINTS; PR00148; ENOLASE.
DR ProDom; PD000902; Enolase; 1.
DR PROSITE; PS00164; ENOLASE; 1.
KW Glycolysis; Lyase; Magnesium.
FT ACT_SITE 322 322 BY SIMILARITY.
FT METAL 334 334 MAGNESIUM (BY SIMILARITY).
FT NON_TER 413 413
SQ SEQUENCE 413 AA; 44404 MW; 11414BCC18644A94 CRC64;

Query Match 37.5%; Score 3; DB 5; Length 413;
Best Local Similarity 100.0%; Pred. No. 5.1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sfxg 8
DB 360 SFXG 363

RESULT 13
Q9TID6
ID Q9TID6 PRELIMINARY; PRT; 475 AA.
AC Q9TID6
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN (EC 4.1.1.39) (RUBISCO
LARGE SUBUNIT).
GN RBCL.
OS Pieris floribunda.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

```

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; Ericales; Ericaceae; Pieris.
OX NCBI_TaxID=49157;
RN [1]
RP SEQUENCE FROM N.A.
RA Kron K.A., Judd W.S., Crayn D.M.;
RT "Phylogenetic analyses of Andromedeae (Ericaceae subfam.
vaccinioideae).";
RL Am. J. Bot. 0:0-0(2000).
CC -1- FUNCTION: RUBISCO CATALYSES TWO REACTIONS: THE CARBOXYLATION OF D-
RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
ACTIVE SITE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) -> 2 3-
PHOSPHO-D-GLYCERATE.
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) -> 3-
PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
DR EMBL; AF124577; AAF16882.1; -.
DR InterPro; IPR000685; Rubisco_large.
DR Pfam; PF00016; RUBISCO_LARGE; 1.
DR PROSITE; PS00157; RUBISCO_LARGE; 1.
KW Carbon dioxide fixation; Chloroplast; Lyase; Monooxygenase;
KW Oxidoreductase; Photorespiration; Photosynthesis.
SQ SEQUENCE 475 AA; 52495 MW; 76F3CA08436481C1 CRC64;

Query Match 37.5%; Score 3; DB 8; Length 475;
Best Local Similarity 100.0%; Pred. No. 5.7e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sfxg 8
DB 61 SFXG 64

RESULT 14
Q9BAX6
ID Q9BAX6 PRELIMINARY; PRT; 510 AA.
AC Q9BAX6
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE MATURASE (FRAGMENT).
GN MATK.
OS Gongora gratulabunda.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Orchidaceae;
OC Epidendroideae; higher Epidendroideae; Maxillarieae; Stanhopeinae;
OC Gongora.
OX NCBI_TaxID=125123;
RN [1]
RP SEQUENCE FROM N.A.
RA Whitten M., Williams N.H., Chase M.W.;
RT "Tribal and subtribal relationships of Maxillarieae (Orchidaceae) with
special emphasis on Stanhopeinae: Combined molecular evidence.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF239478; AAK31875.1; -.
KW Chloroplast.
FT NON_TER 1 1
FT NON_TER 510 510
SQ SEQUENCE 510 AA; 61288 MW; C70C056B44F26B8E CRC64;

Query Match 37.5%; Score 3; DB 8; Length 510;
Best Local Similarity 100.0%; Pred. No. 6e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 5 sfxxg 8
|||||
Db 346 SFXG 349

RESULT 15
Q9BAX0 PRELIMINARY; PRT; 510 AA.
AC Q9BAX0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE MATURASE (FRAGMENT).
GN MATK.
OS
GN Gongora sphaerica.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Orchidaceae;
OC Epidendroideae; higher Epidendroideae; Maxillariaceae; Stanhopeinae;
OC Gongora.
OX NCBI_TaxID=125126;
RN [1]
SEQUENCE FROM N.A.
RP Whitten M., Williams N.H., Chase M.W.;
RA "Tribal and subtribal relationships of Maxillariaceae (Orchidaceae) with
RT special emphasis on Stanhopeinae: Combined molecular evidence.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF239484; AK31881.1; -;
KW Chloroplast.
FT NON_TER 1 1
FT NON_TER 510 510
SQ SEQUENCE 510 AA; 61361 MW; B03CD0D32738AAB3 CRC64;

Query Match 37.5%; Score 3; DB 8; Length 510;
Best Local Similarity 100.0%; Pred. No. 6e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 sfxxg 8
|||||
Db 346 SFXG 349

RESULT 16
Q32665 PRELIMINARY; PRT; 698 AA.
ID Q32665;
AC Q32665;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT (FRAGMENT).
GN NDHF.
OS Nicotinia diversifolia.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Lamiales; Scrophulariaceae; Nicotemia.
OX NCBI_TaxID=28500;
RN [1]
SEQUENCE FROM N.A.
RP Olmstead R.G., Reeves P.A.;
RA Ann. Mo. Bot. Gard. 82:176-193(1995).
RL -!- CATALYTIC ACTIVITY: NADH + PLASTOQUINONE = NAD(+) + PLASTOQUINOL.
CC -!- SIMILARITY: TO NADH-UBIQUINONE/PLASTOQUINONE (COMPLEX I), VARIOUS
CC CHAINS
DR EMBL; L36405; AAA84496.1; -;
DR Mendel; 2471; Nicd1;ndhf;2471.
DR InterPro; IPR001064; Crystallin.
DR InterPro; IPR001750; Oxidored_q1.
DR InterPro; IPR002128; Oxidored_q1.C.
DR InterPro; IPR001516; Oxidored_q1_N.
DR Pfam; PF00361; oxidored_q1; 1.
DR Pfam; PF01010; oxidored_q1_C; 1.

DR Pfam; PF00662; oxidored_q1_N; 1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
KW Chloroplast; NAD; Oxidoreductase; Plastoquinone.
FT NON_TER 1 1
FT NON_TER 698 698
SQ SEQUENCE 698 AA; 79217 MW; F4B4EBE4E7440A3D CRC64;

Query Match 37.5%; Score 3; DB 8; Length 698;
Best Local Similarity 100.0%; Pred. No. 7.6e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 sfxxg 8
|||||
Db 664 SFXG 667

RESULT 17
Q9MVF7 PRELIMINARY; PRT; 707 AA.
ID Q9MVF7;
AC Q9MVF7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE NADH DEHYDROGENASE (FRAGMENT).
GN NDHF.
OS Muntingia calabura.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OX eurosids II; Malvales; Muntingiaceae; Muntingia.
OX NCBI_TaxID=45164;
RN [1]
SEQUENCE FROM N.A.
RP Alverson W.S., Whitlock B.A., Nyffeler R., Bayer C., Baum D.A.;
RA "Phylogeny of the core Malvales: evidence from ndhF sequence data.";
RL Am. J. Bot. 86:1459-1471(1999).
CC -!- CATALYTIC ACTIVITY: NADH + PLASTOQUINONE = NAD(+) + PLASTOQUINOL.
CC -!- SIMILARITY: TO NADH-UBIQUINONE/PLASTOQUINONE (COMPLEX I), VARIOUS
CC CHAINS
DR EMBL; AF111781; AAF27257.1; -;
DR InterPro; IPR001750; Oxidored_q1.
DR InterPro; IPR002128; Oxidored_q1_C.
DR InterPro; IPR001516; Oxidored_q1_N.
DR Pfam; PF00361; oxidored_q1; 1.
DR Pfam; PF01010; oxidored_q1_C; 1.
DR Pfam; PF00662; oxidored_q1_N; 1.
KW Chloroplast; NAD; Oxidoreductase; Plastoquinone.
FT NON_TER 707 707
SQ SEQUENCE 707 AA; 79937 MW; 9D439E8899FF6393 CRC64;

Query Match 37.5%; Score 3; DB 8; Length 707;
Best Local Similarity 100.0%; Pred. No. 7.7e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 kxxsf 6
|||||
Db 503 KXXSF 507

RESULT 18
Q9MVG6 PRELIMINARY; PRT; 717 AA.
ID Q9MVG6;
AC Q9MVG6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE NADH DEHYDROGENASE (FRAGMENT).
GN NDHF.
OS Luehea seemannii.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Malvales; Malvaceae; Luehea.
 OX NCBI_TaxID=45194;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Alverson W.S., Whitlock B.A., Nyffeler R., Bayer C., Baum D.A.;
 RT "Phylogeny of the core Malvales: evidence from ndhF sequence data."
 RL Am. J. Bot. 86:1459-1471(1999).
 CC -!- CATALYTIC ACTIVITY: NADH + PLASTOQUINONE = NAD(+) + PLASTOQUINOL.
 CC -!- SIMILARITY: TO NADH-UBIQUINONE/PLASTOQUINONE (COMPLEX I), VARIOUS
 CC CHAINS.
 DR EMBL: AF111770; AAF27248.1; -.
 DR InterPro: IPR001750; Oxidored_q1.
 DR InterPro: IPR002128; Oxidored_q1_C.
 DR InterPro: IPR001516; Oxidored_q1_N.
 DR Pfam: PF00361; Oxidored_q1.
 DR Pfam: PF01010; Oxidored_q1_C; 1.
 DR Pfam: PF00662; Oxidored_q1_N; 1.
 DR Chloroplast; NAD; Oxidoreductase; Plastoquinone.
 KW NON_TER 717 717
 FT SEQUENCE 717 AA; 81293 MW; DF845EF5B33D64A7 CRC64;
 SQ SEQUENCE 717 AA; 81293 MW; DF845EF5B33D64A7 CRC64;

Query Match 37.5%; Score 3; DB 8; Length 717;
 Best Local Similarity 100.0%; Pred. No. 7.8e+03;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 kxxsf 6
 |||||
 Db 469 KXSF 473

RESULT 19
 O07354 PRELIMINARY; PRT; 7 AA.
 ID O07354
 AC O07354;
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
 DE NIFK (FRAGMENT).
 GN NIFK.
 OS Synechococcus sp. (strain PCC 8801 / RF-1) (Cyanothecae PCC 8801).
 OC Bacteria; Cyanobacteria; Chroococcales; Cyanothecae.
 OX NCBI_TaxID=41431;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RF-1.
 RA Chen H.M.;
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF003700; AAC35193.1; -.
 FT NON_TER 1 1
 SQ SEQUENCE 7 AA; 849 MW; 7412C72AA9D5B030 CRC64;

Query Match 25.0%; Score 2; DB 2; Length 7;
 Best Local Similarity 100.0%; Pred. No. 4.7e+05;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sf 6
 ||
 Db 2 SF 3

RESULT 20
 Q95945 PRELIMINARY; PRT; 7 AA.
 ID Q95945
 AC Q95945;
 DT 01-FEB-1997 (TrEMBLrel. 02, Created)
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
 DT 01-FEB-1998 (TrEMBLrel. 08, Last annotation update)
 DE INSIDE INTRON 5 (FRAGMENT).
 OS Saccharomyces cerevisiae (Baker's yeast).
 OG Mitochondrion.

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=81069885; PubMed=6254986;
 RT Bonitz S.G., Coruzzi G., Thalenfeld B., Tzagoloff A., Macino G.;
 RT "Assembly of the mitochondrial membrane system. Structure and
 RT nucleotide sequence of the gene coding for subunit 1 of yeast
 RT cytochrome oxidase."
 RL J. Biol. Chem. 255:11927-11941(1980).
 DR EMBL: V00694; CAA24066.1; -.
 KW Mitochondrion.
 FT NON_TER 1 1
 SQ SEQUENCE 7 AA; 859 MW; 75B7232362CDC460 CRC64;

Query Match 25.0%; Score 2; DB 8; Length 7;
 Best Local Similarity 100.0%; Pred. No. 4.7e+05;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wk 2
 ||
 Db 4 WK 5

RESULT 21
 P87225 PRELIMINARY; PRT; 8 AA.
 ID P87225
 AC P87225;
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE GIN11 PROTEIN (FRAGMENT).
 GN GIN11 OR YLL065W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wedler H., Wedler E., Scharfe M., Wambutt R.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MIPS;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z73169; CAA97518.2; -.
 DR SGD: S000398; GIN11.
 FT NON_TER 1 1
 SQ SEQUENCE 8 AA; 1019 MW; 4E21A9C449D5B73B CRC64;

Query Match 25.0%; Score 2; DB 3; Length 8;
 Best Local Similarity 100.0%; Pred. No. 4.7e+05;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sf 6
 ||
 Db 3 SF 4

RESULT 22
 Q9HCQ0 PRELIMINARY; PRT; 8 AA.
 ID Q9HCQ0
 AC Q9HCQ0;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE PHOSPHODIESTERASE 10A7 (PDE10A7) (FRAGMENT).
 GN HSPDE10A.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=10998054;
RA Fujishige K., Kotera J., Yuasa K., Omori K.;
RT "The human phosphodiesterase PDE10A gene. Genomic organization and
RT evolutionary relatedness with other PDEs containing GAF domains.";
RL Eur. J. Biochem. 267:5943-5951(2000).
DR EMBL; AB041779; BAB16368.1; -;
FT NON_TER 8
SQ SEQUENCE 8 AA; 966 MW; FD4B19D5A6C76446 CRC64;

Query Match 25.0%; Score 2; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 sf 6
||
Db 5 SF 6

RESULT 23
P82685
ID P82685 PRELIMINARY; PRT; 8 AA.
AC P82685;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE KININ-1 (PEA-K-1).
OS Periplaneta americana (American cockroach).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC Blattoidea; Blattidae; Periplaneta.
OX NCBI_TaxID=6978;
RN [1]
RP SEQUENCE, AND FUNCTION.
RC TISSUE=CORPORA CARDIACA;
RX MEDLINE=98010462; PubMed=9350979;
RA "Isolation and structural elucidation of eight kinins from the
RT retrocerebral complex of the American cockroach, Periplaneta
RT americana.";
RL Regul. Pept. 71:199-205(1997).
CC -!- FUNCTION: MEDIATES VISCERAL MUSCLE CONTRACTILE ACTIVITY
CC -!- (MYOTROPIC ACTIVITY).
CC -!- SIMILARITY: BELONGS TO THE KININ FAMILY.
KW Neuropeptide; Amidation.
FT MOD_RES 8
SQ SEQUENCE 8 AA; 950 MW; 326365B449D5A774 CRC64;

Query Match 25.0%; Score 2; DB 5; Length 8;
Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 sf 6
||
Db 3 SF 4

RESULT 24
P82686
ID P82686 PRELIMINARY; PRT; 8 AA.
AC P82686;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE KININ-2 (PEA-K-2).
OS Periplaneta americana (American cockroach).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;

OC Blattoidea; Blattidae; Periplaneta.
OX NCBI_TaxID=6978;
RN [1]
RP SEQUENCE, AND FUNCTION.
RX TISSUE=CORPORA CARDIACA;
RX MEDLINE=98010462; PubMed=9350979;
RA Predel R., Kellner R., Rapus J., Penzlin H., Gaede G.;
RT "Isolation and structural elucidation of eight kinins from the
RT retrocerebral complex of the American cockroach, Periplaneta
RT americana.";
RL Regul. Pept. 71:199-205(1997).
CC -!- FUNCTION: MEDIATES VISCERAL MUSCLE CONTRACTILE ACTIVITY
CC -!- (MYOTROPIC ACTIVITY).
CC -!- SIMILARITY: BELONGS TO THE KININ FAMILY.
KW Neuropeptide; Amidation.
FT MOD_RES 8
SQ SEQUENCE 8 AA; 856 MW; DC6365A5B9D5BDDA CRC64;

Query Match 25.0%; Score 2; DB 5; Length 8;
Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 sf 6
||
Db 3 SF 4

RESULT 25
P82687
ID P82687 PRELIMINARY; PRT; 8 AA.
AC P82687;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE KININ-3 (PEA-K-3).
OS Periplaneta americana (American cockroach).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC Blattoidea; Blattidae; Periplaneta.
OX NCBI_TaxID=6978;
RN [1]
RP SEQUENCE, AND FUNCTION.
RC TISSUE=CORPORA CARDIACA;
RX MEDLINE=98010462; PubMed=9350979;
RA Predel R., Kellner R., Rapus J., Penzlin H., Gaede G.;
RT "Isolation and structural elucidation of eight kinins from the
RT retrocerebral complex of the American cockroach, Periplaneta
RT americana.";
RL Regul. Pept. 71:199-205(1997).
CC -!- FUNCTION: MEDIATES VISCERAL MUSCLE CONTRACTILE ACTIVITY
CC -!- (MYOTROPIC ACTIVITY).
CC -!- SIMILARITY: BELONGS TO THE KININ FAMILY.
KW Neuropeptide; Amidation.
FT MOD_RES 8
SQ SEQUENCE 8 AA; 909 MW; DC6365B449D5A76A CRC64;

Query Match 25.0%; Score 2; DB 5; Length 8;
Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 sf 6
||
Db 3 SF 4

Search completed: January 14, 2002, 08:07:33
Job time: 763 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:56:29 ; Search time 81.39 Seconds
(without alignments)
7.281 Million cell updates/sec

Title: 09-185908-1C
Perfect score: 8
Sequence: 1 wkxxsfxxg 8

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 50 summaries

Database : A_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3	37.5	31	AAW62083	Hyphozyma sp. stra
2	3	37.5	40	AAV40039	Peptide sequence d
3	3	37.5	42	AAW77559	Staphylococcus aur
4	3	37.5	58	AAV12820	Human 5' EST secre
5	3	37.5	84	AAV76222	Human colon cancer
6	3	37.5	114	AAW40785	Human polypeptide
7	3	37.5	115	AAW21195	Exol8 partial prot
8	3	37.5	129	AAW20281	H. pylori cytoplas
9	3	37.5	129	AAW24625	H. pylori cytoplas
10	3	37.5	149	AAW53916	Human colon cancer
11	3	37.5	156	AAW56254	Interleukin-13 bin

12	3	37.5	174	22	AAW63508	Human gastric canc
13	3	37.5	194	18	AAW28289	Amino acid sequenc
14	3	37.5	217	22	AAW63791	Human prostate can
15	3	37.5	229	22	AAW72280	Human olfactory re
16	3	37.5	297	22	AAW41776	Human polypeptide
17	3	37.5	312	22	AAW72750	Human olfactory re
18	3	37.5	312	22	AAW72752	Human olfactory re
19	3	37.5	354	21	AAW76127	Human secreted pro
20	3	37.5	500	19	AAW48932	Human papillomavir
21	3	37.5	552	22	AAW48242	Amino acid sequenc
22	3	37.5	908	16	AAW85443	Rat brevicin core
23	3	37.5	1186	20	AAW58339	Banana ripening fr
24	3	37.5	2466	20	AAW05842	Banana ripening fr
25	2	25.0	3	15	AAW42556	ACE inhibitor SP3.
26	2	25.0	3	21	AAW37940	Trypsin modulating
27	2	25.0	4	6	AAW50545	Sequence of Cyclo(
28	2	25.0	4	7	AAW60801	Peptide with neutr
29	2	25.0	4	8	AAW71285	Opiate binding pep
30	2	25.0	4	8	AAW71288	Opiate binding pep
31	2	25.0	4	9	AAW81591	Organ specific neo
32	2	25.0	4	13	AAW20049	Tetrapeptide chole
33	2	25.0	4	13	AAW24950	Conformationally c
34	2	25.0	4	13	AAW30195	Des-Tyr(1)-beta-ca
35	2	25.0	4	13	AAW30202	Des-Tyr(1)-beta-ca
36	2	25.0	4	13	AAW30206	Des-Tyr(1)-beta-ca
37	2	25.0	4	14	AAW38135	Protease-non-labill
38	2	25.0	4	14	AAW38130	Protease-non-labill
39	2	25.0	4	14	AAW38132	Protease-non-labill
40	2	25.0	4	14	AAW38113	Protease-non-labill
41	2	25.0	4	14	AAW38115	Protease-non-labill
42	2	25.0	4	15	AAW57101	Epitope P2 of anti
43	2	25.0	4	15	AAW46007	Serine protease in
44	2	25.0	4	15	AAW46009	Serine protease in
45	2	25.0	4	15	AAW46005	Serine protease in
46	2	25.0	4	15	AAW46003	Serine protease in
47	2	25.0	4	15	AAW46008	Serine protease in
48	2	25.0	4	15	AAW46006	Serine protease in
49	2	25.0	4	15	AAW46004	Serine protease in
50	2	25.0	4	15	AAW50277	Somatostatin pepti

ALIGNMENTS

RESULT 1

AAW62083
ID AAW62083 standard; peptide; 31 AA.

XX AC AAW62083;

XX DT 15-SEP-1998 (first entry)

XX DE Hyphozyma sp. strain CBS 648.91 phospholipase fragment #8.

XX KW Hyphozyma sp. strains CBS 648.91; phospholipase; fatty acyl; hydrolyase;
XX KW oil degumming; wheat starch hydrolysate; breadmaking; dough.

XX OS Hyphozyma sp.

XX FH Key Location/Qualifiers

FT Misc-difference 5 /note= "believed to be a glycosylated Asn residue"

FT FT

XX WO9818912-A1.

XX PD 07-MAY-1998.

XX PF 30-OCT-1997; 97WO-DK00490.

XX PR 31-OCT-1996; 96DK-0001215.

XX XX (NOVO) NOVO-NORDISK AS.

XX PA

PI Halkier T, Hasiida M, Stringer MA, Tsutsumi N;
XX WPI; 1998-272208/24.
XX
XX New phospholipase from *Hyphozyma* strain and related DNA - hydrolyses
PT both acyl residues in phospholipid, useful for degumming edible
PT oils, and to improve dough quality or filterability of carbohydrate
PT slurries
XX
XX Claim 3; Page 18; 46pp; English.
XX
XX The present sequence represents a new phospholipase peptide fragment,
CC from strain CBS 648.91 of a *Hyphozyma*. The phospholipase can hydrolyse
CC both fatty acyl groups in a phospholipid (PL), has an optimum temperature
CC about 50 degrees Celsius (measured for 10 min at pH 3-4) and optimum pH
CC about 3 (measured at 40 degrees Celsius for 10 min). The phospholipase is
CC used to hydrolyse acyl groups in (lyso)PL, particularly (lyso)lecithin.
CC Specifically it is used: (a) to improve filterability of aqueous
CC solutions or slurries of carbohydrate origin (especially wheat starch
CC hydrolysate) containing PL; (b) in breadmaking (added to the dough) to
CC improve elasticity; and (c) to reduce PL content (degum) in edible oils.
CC The phospholipase lacks lipase activity and is effective at very low pH.
CC so does not cause enzymatic/alkaline hydrolysis of triglycerides. It is
CC not membrane bound so can be produced and purified on a commercial
CC scale.
XX
XX Sequence 31 AA;

Query Match 37.5%; Score 3; DB 19; Length 31;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sfxg 8
| | | |
Db 3 sfxg 6

RESULT 2
AAV40039
ID AAV40039 standard; Peptide; 40 AA.
XX
XX AC AAV40039;
XX
XX 18-NOV-1999 (first entry)
XX
XX Peptide sequence derived from a human secreted protein.

XX Secreted protein; gene therapy; cancer; tumor; fetal deficiency;
KW neurodegenerative disorder; developmental abnormality; blood disorder;
KW immune system disease; autoimmune disease; leukemia; inflammation;
KW allergy; Alzheimer's disease; cognitive disorder; schizophrenia;
KW obesity; osteoporosis; arthritis; infection; AIDS; diabetes; asthma;
KW connective tissue disorder; transplant rejection; sepsis; acne;
KW psoriasis; cardiovascular disorder; reproductive disorder;
KW food additive; food preservative; storage capability.

XX Homo sapiens.
XX
XX WO9943693-A1.
XX
XX 02-SEP-1999.
XX
XX 24-FEB-1999; 99WO-US03939.
XX
XX 26-FEB-1998; 98US-0076051.
PR 26-FEB-1998; 98US-0076052.
PR 26-FEB-1998; 98US-0076053.
PR 26-FEB-1998; 98US-0076054.
PR 26-FEB-1998; 98US-0076057.
XX
XX (HUMA-) HUMAN GENOME SCI INC.

PI Olsen HS, Florence K, Brewer LA, Ebner R, Ruben SM, Rosen CA;
XX Duan RD;
XX
XX WPI; 1999-550857/46.
XX
XX New human genes and the secreted polypeptides they encode, useful for
PT diagnosis and treatment of e.g. cancers, neurological disorders, immune
PT diseases, inflammation or blood disorders
XX
XX Disclosure; Page 39; 246pp; English.

XX
XX AAV40001-92 are derived from human secreted proteins. The
CC polynucleotides and their corresponding secreted polypeptides are useful
CC for preventing, treating or ameliorating medical conditions, e.g. by
CC protein or gene therapy. Pathological conditions can also be diagnosed by
CC determining the amount of the new polypeptides in a sample or by
CC determining the presence of mutations in the polynucleotide. Specific
CC uses include developing products for the diagnosis or treatment of
CC cancer, tumors, neurodegenerative disorders, developmental abnormalities
CC and fetal deficiencies, blood disorders, sepsis, diseases of the immune
CC system, autoimmune diseases, inflammation, allergies, Alzheimer's and
CC cognitive disorders, schizophrenia, obesity, osteoporosis, arthritis,
CC infections, AIDS, connective tissue disorders, transplant rejection,
CC diabetes, asthma, sepsis, acne, psoriasis, cardiovascular disorders,
CC and reproductive disorders. The polypeptides or polynucleotides can
CC also be used as food additives or preservatives, such as to increase
CC or decrease storage capabilities, fat content, lipid, protein,
CC carbohydrate, vitamins, minerals, cofactors or other nutritional
CC components.

XX
XX Sequence 40 AA;

Query Match 37.5%; Score 3; DB 20; Length 40;
Best Local Similarity 100.0%; Pred. No. 4.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sfxg 8
| | | |
Db 7 sfxg 10

RESULT 3
AAW77559
ID AAW77559 standard; Protein; 42 AA.

XX
XX AC AAW77559;
XX
XX 30-OCT-1998 (first entry)
XX
XX Staphylococcus aureus protein of unknown function.

XX Staphylococcus aureus protein; immune response induction; eye infection;
KW antibody production; T-cell immune response; gastrointestinal infection;
KW respiratory infection; inhibitor; bacterial infection; cardiac infection;
KW central nervous system; kidney infection; urinary tract infection;
KW antimicrobial compound identification; broad spectrum antibiotic;
KW therapy.

XX Staphylococcus aureus.

XX
XX Key Location/Qualifiers
FT Misc-difference 1..42
FT /note= "residues designated X are unspecified, and
FT represented as Xaa in the specification"

XX EP841394-A2.
XX
XX 13-MAY-1998.
XX
XX 24-SEP-1997; 97EP-0307485.
XX
XX 24-SEP-1996; 96US-0027032.


```
XX PA (SMIK ) SMITHKLINE BEECHAM CORP.
XX PA (SMIK ) SMITHKLINE BEECHAM PLC.
XX PI Black MT, Burnham MKR, Hodgson JE, Knowles DJC;
XX PI Lonetto MA, Nicholas RO, Pratt JM, Reichard RW, Rosenberg M;
XX PI Ward JM;
XX DR WPI; 1998-252940/23.
XX DR N-PSDB; AAV53359.
XX XX
XX PT New nucleic acid sequences from Staphylococcus aureus WCHU29 -
XX PT useful in vaccines and for treatment of bacterial infections of e.g.
XX PT respiratory tract and central nervous system
XX XX
XX PS Claim 11; Page 252; 390pp; English.
XX CC This sequence represents a Staphylococcus aureus protein of unknown
XX CC function, and is encoded by a DNA sequence of the invention.
XX CC The DNA sequences were isolated from Staphylococcus aureus WCHU29
XX CC (NCIMB 40771). Host cells containing the DNA sequences are used to
XX CC produce polypeptides or fragments. The proteins are used in the treatment
XX CC of disease, for inducing an immune response by administering them, to
XX CC produce antibody and/or T-cell immune response. Antagonists of the
XX CC proteins are used for the inhibition of bacterial polypeptides.
XX CC Conditions which may be treated include bacterial infections, especially
XX CC respiratory, cardiac, gastrointestinal, central nervous, eye, kidney,
XX CC urinary tract, skin, bones and joints. The proteins can also be used to
XX CC identify antimicrobial compounds which are broad spectrum antibiotics,
XX CC especially useful in the treatment of H. pylori infection.
XX XX
XX SQ Sequence 42 AA;
Query Match 37.5%; Score 3; DB 19; Length 42;
Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 sfxg 8
Db 34 sfxg 37
|||||
RESULT 4
AAV12820
ID AAV12820 standard; Protein; 58 AA.
XX AC AAV12820;
XX DT 21-JUN-1999 (first entry)
XX DE Human 5' EST secreted protein SEQ ID NO:410.
XX KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
XX KW forensic; gene therapy; chromosome mapping; signal peptide;
XX KW upstream regulatory sequence; cytokine activity; cell proliferation;
XX KW differentiation; haematopoiesis regulation; tissue growth regulation;
XX KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
XX KW thrombolytic; anti-inflammatory; tumour inhibition.
XX XX
XX OS Homo sapiens.
XX XX WO9906549-A2.
XX PN 11-FEB-1999.
XX PD 31-JUL-1998; 98WO-1B01231.
XX PF 01-AUG-1997; 97US-0905279.
XX PR (GEST ) GENSET.
XX PA Duclert A, Dumas Milne Edwards J, Lacroix B;
XX PI
```

```
XX DR WPI; 1999-153779/13.
XX DR N-PSDB; AAX51598.
XX XX
XX PT New nucleic acids encoding human secreted proteins - obtained from
XX PT CDNA libraries derived from testis, ovary, uterus and spleen tissue
XX PS Claim 34; Page 465; 522pp; English.
XX CC AAX51459 to AAX51691 represent 5' expressed sequence tags (ESTs) for
XX CC human secreted proteins, and encode the proteins given in AAV12681 to
XX CC AAV12913, respectively. The proteins given represent the signal peptide
XX CC and an N-terminal fragment of a secreted protein. The nucleic acid
XX CC sequences can be used for producing secreted human gene products. They
XX CC can also be used to develop products for diagnosis and therapy. The
XX CC proteins obtained may have cytokine activity, cell
XX CC proliferation/differentiation activity, haematopoiesis regulating
XX CC activity, tissue growth regulating activity, reproductive hormone
XX CC regulating activity, chemotactic/ chemokinetic activity, haemostatic and
XX CC thrombolytic activity, receptor/ ligand activity, anti-inflammatory
XX CC activity, tumour inhibition activity or other activities. The products
XX CC can be used in forensic, gene therapy and chromosome mapping procedures.
XX CC The sequences can also be used for obtaining corresponding promoter
XX CC sequences. The nucleic acids encoding the signal peptide can be used for
XX CC directing extracellular secretion of a polypeptide or the insertion of a
XX CC polypeptide into a membrane, or importing a polypeptide into a cell.
XX XX
XX SQ Sequence 58 AA;
Query Match 37.5%; Score 3; DB 20; Length 58;
Best Local Similarity 100.0%; Pred. No. 6.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 kxxsf 6
Db 23 kxxsf 27
|||||
RESULT 5
AAG76222
ID AAG76222 standard; Protein; 84 AA.
XX AC AAG76222;
XX DT 03-SEP-2001 (first entry)
XX DE Human colon cancer antigen protein SEQ ID NO:6986.
XX KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX KW colorectal carcinoma.
XX XX Homo sapiens.
XX OS WO200122920-A2.
XX PN 05-APR-2001.
XX PD 28-SEP-2000; 2000WO-US26524.
XX PF 29-SEP-1999; 99US-0157137.
XX PR 03-NOV-1999; 99US-0163280.
XX XX (HUMA-) HUMAN GENOME SCI INC.
XX XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX PI WPI; 2001-235357/24.
XX DR N-PSDB; AAH35627.
XX XX
XX PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX XX
```

PS Claim 11; Page 8423-8426; 9803pp; English.
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAH77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 84 AA;

Query Match 37.5%; Score 3; DB 22; Length 84;
Best Local Similarity 100.0%; Pred. No. 7.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sfxg 8
Db 28 sfxg 31

RESULT 6
AAH40785
ID AAM40785 standard; Protein; 114 AA.
XX
AC AAM40785;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 5716.

XX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 23-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.
DR N-PSDB; AAI59941.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 2; SEQ ID NO 5716; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 114 AA;

Query Match 37.5%; Score 3; DB 22; Length 114;
Best Local Similarity 100.0%; Pred. No. 9.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sfxg 8
Db 78 sfxg 81

RESULT 7
AAB21195
ID AAB21195 standard; Protein; 115 AA.
XX
AC AAB21195;
XX
DT 12-JAN-2001 (first entry)
XX
DE Exol8 partial protein.
XX
KW Exocytosis pathway protein; Exo protein; antiparkinsonian; antidiabetic;
KW antiallergic; antiasthmatic; neurotropic; neuroprotective; anticonvulsant;
KW vulneryary; asthma; inflammation; allergy; Chediak-Higashi syndrome; CHS;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease; diabetes;
KW digestion disorder; wound healing disorder; gene therapy.
XX
OS Mus sp.
XX
PN WO2000043419-A2.
XX
PD 27-JUL-2000.
XX
PF 20-JAN-2000; 2000WO-US01431.
XX
PR 20-JAN-1999; 99US-0116534.
PR 26-JAN-1999; 99US-0117274.
PR 26-JAN-1999; 99US-0117308.
PR 26-JAN-1999; 99US-0117309.
PR 26-JAN-1999; 99US-0117312.
PR 01-FEB-1999; 99US-0118177.
PR 01-FEB-1999; 99US-0118178.
PR 01-FEB-1999; 99US-0118179.
PR 09-FEB-1999; 99US-0119286.
PR 11-FEB-1999; 99US-0119998.
PR 11-FEB-1999; 99US-0119759.

XX (RICE-) RIGEL PHARM INC.
XX
XX
XX Luo Y;
XX
XX WPI: 2000-482908/42.
XX N-PSDB; AAA89574.
XX
XX New nucleic acids encoding Exo proteins which are useful in the
XX diagnosis, treatment or prevention of exocytosis-mediated disorders
XX such as asthma, inflammation and allergies -
XX
XX Disclosure; Page 167; 305pp; English.
XX
XX The present sequence is a polypeptide which is associated with
XX the exocytosis pathway. cDNA molecules encoding proteins involved in
XX exocytosis have been isolated by yeast one-hybrid and two-hybrid
XX screening. Novel proteins, termed Exo proteins, have been identified that
XX interact with known exocytosis-associated proteins such as GS27, alpha
XX snap, unc18-1, vamps3, snap23, and the rab family of proteins.
XX Exo proteins and their agonists and antagonists are useful in the
XX diagnosis, treatment or prevention of exocytosis-mediated disorders
XX such as asthma, inflammation, allergies, Chediak-Higashi Syndrome
XX (CHS), Alzheimer's disease, Parkinson's disease, Huntington's disease,
XX diabetes, digestion disorders and wound healing disorders.
XX The nucleic acids, antagonists or agonists of Exo proteins are useful
XX in gene therapy. The nucleic acids are also useful for generating
XX transgenic or knock-out animals which can be used in the
XX development and screening of therapeutically useful reagents.
XX
XX Sequence 115 AA;
SQ

Query Match 37.5%; Score 3; DB 21; Length 115;
Best Local Similarity 100.0%; Pred. No. 9.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sfxxg 8
||||
Db 107 sfxxg 110

RESULT 8
AAW20281
ID AAW20281 standard; Protein: 129 AA.
XX
XX AAW20281;
XX
XX 30-JUL-1997 (first entry)
XX
XX H. pylori cytoplasmic protein 24039587.aa.
XX
XX Cytoplasmic; vaccine; prevention; treatment; infection; identification;
XX binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
XX duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
XX
XX Helicobacter pylori.
XX
XX Key Location/Qualifiers
XX Misc-difference 115
XX /label= unknown
XX /note= "encoded by AAW"
XX
XX WO9640893-A1.
XX
XX 19-DEC-1996.
XX
XX 06-JUN-1996; 96WO-US09122.
XX
XX 01-APR-1996; 96US-0630405.
XX 07-JUN-1995; 95US-0487032.
XX
XX (ASTR) ASTRA AB.
XX

XX Berglindh OT, Smith D, Mellgaard BL;
XX
XX WPI: 1997-052306/05.
XX N-PSDB; AAT67763.
XX
XX Helicobacter pylori nucleic acid sequences and related
XX polypeptides) - useful for vaccines to treat or prevent H. pylori
XX infection, and to detect Helicobacter
XX
XX Claim 61; Page 482; 1481pp; English.
XX
XX This sequence is a H. pylori cytoplasmic protein.
XX The protein may be used in a vaccine to prevent or treat H. pylori
XX infection or to identify H. pylori polypeptide binding compounds,
XX useful as potential H. pylori life cycle activators or inhibitors.
XX The genomic sequence of H. pylori (ATCC 55679) was determined from
XX overlapping contigs generated by mechanically shearing the bacterial
XX DNA. The sequences were analysed for ORF of at least 180 nucleotides,
XX and the predicted coding regions defined by computer evaluation. To
XX identify likely H. pylori antigens for vaccine development, the amino
XX acid sequences predicted from various ORF were analysed for significant
XX homology to other known or exported membrane proteins. Having identified
XX and determined the sequences of interest, particular regions can be
XX isolated from H. pylori by PCR amplification for recombinant polypeptide
XX production, e.g. in E. coli hosts.
XX
XX Sequence 129 AA;
SQ

Query Match 37.5%; Score 3; DB 18; Length 129;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sfxxg 8
||||
Db 113 sfxxg 116

RESULT 9
AAW24625
ID AAW24625 standard; Protein: 129 AA.
XX
XX AAW24625;
XX
XX 11-AUG-1997 (first entry)
XX
XX H. pylori cytoplasmic protein 24039587.aa.
XX
XX Transmembrane; cytoplasmic; cell envelope; flagella; transport;
XX secreted; periplasmic; chronic gastritis; duodenal ulcer disease;
XX activator; inhibitor; bacterial life cycle; vaccine; immunise;
XX detection; antisense; inhibition.
XX
XX Helicobacter pylori.
XX
XX Key Location/Qualifiers
XX Misc-difference 115
XX /note= "encoded by AAW"
XX
XX WO9719098-A1.
XX
XX 29-MAY-1997.
XX
XX 15-NOV-1996; 96WO-US18542.
XX
XX 17-NOV-1995; 95US-0561469.
XX
XX (ASTR) ASTRA AB.
XX
XX Smith DH;
XX
XX WPI: 1997-298052/27.
XX

DR N-PSDB; AAT77443.
XX Helicobacter pylori nucleic acid sequences and related proteins -
PT used for diagnostics and therapeutics
XX
PS Claim 10; Page 129; 1481pp; English.
XX
CC This sequence is a H. pylori cytoplasmic protein.
CC Helicobacter pylori has been strongly linked to chronic gastritis and
CC duodenal ulcer disease. The nucleic acid sequences of the invention
CC are used to evaluate compounds, especially activators or inhibitors of
CC bacterial life cycle, for the ability to bind an H. pylori nucleic acid
CC sequence. The nucleic acid sequences, and corresponding proteins, are
CC also useful for generating vaccines for immunising subjects against H.
CC pylori or for use in detecting the presence of Helicobacter species in
CC a sample. Antisense nucleic acid sequences of these sequences are
CC used to inhibit expression of a gene from Helicobacter species. H.
CC pylori whole genomic DNA was isolated and nebulised to a median size of
CC 2000 bp. Purified DNA fragments were blunt-ended and ligated to unique
CC BstXI-linker adapters in 100-1000 fold molar excess. These linkers are
CC complementary to the BstXI-cut PMPX vectors, while the overhang is not
CC self-complementary. Therefore the linkers will not concatamerise nor
CC will the cut vector re-ligate itself easily. The linker-adaptor inserts
CC were ligated to each of the 20 PMPX vectors to construct a series of
CC shotgun subclone libraries. The purified DNA samples were then
CC sequenced.
CC Note: The ORF/protein reference number for this sequence was obtained
CC from the related specification, WO9640893.
XX
SQ Sequence 129 AA;

Query Match 37.5%; Score 3; DB 18; Length 129;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sfxg 8
Db 113 sfxg 116
|||||

RESULT 10
AAB53916
ID AAB53916 standard; Protein; 149 AA.
XX AC AAB53916;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human colon cancer antigen protein sequence SEQ ID NO:1456.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytostatic; cardioactive; neuroprotective; vulnary;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder.
XX
OS Homo sapiens.
XX
PN WO200055351-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05883.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX

DR WPI; 2000-587534/55.
DR N-PSDB; AAC98673.
XX
PT Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer -
XX
PS Claim 11; Page 2017; 2104pp; English.
XX
CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnary, nephrotropic, antiinfective and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, wounds, renal disorders, infectious
CC gastrointestinal disorders, cardiovascular disorders. AAC98764 to AAC98772 and
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 149 AA;

Query Match 37.5%; Score 3; DB 21; Length 149;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sfxg 8
Db 135 sfxg 138
|||||

RESULT 11
AAW56254
ID AAW56254 standard; Protein; 156 AA.
XX AC AAW56254;
XX
DT 16-SEP-1998 (first entry)
XX
DE Interleukin-13 binding protein N-terminal ORF 3.
XX
KW Therapeutic; IL-3 mediated condition; allergy; asthma; diagnosis;
KW autoimmune disease; antibody; immunotherapy.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 7 /label= Unknown
FT /note= "Encoded by nct"
FT Misc-difference 20 /note= "Stop codon"
FT /note= "Stop codon"
FT Misc-difference 74 /note= "Stop codon"
FT /note= "Stop codon"
FT Misc-difference 104 /note= "Stop codon"
FT /note= "Stop codon"
FT Misc-difference 111 /note= "Stop codon"
FT /note= "Stop codon"
FT Misc-difference 119 /note= "Stop codon"
FT /note= "Stop codon"
FT Misc-difference 120 /note= "Stop codon"
FT /note= "Stop codon"
FT Misc-difference 140 /label= Unknown
FT /note= "Encoded by tcn"
XX

PN WO9810638-A1.
XX 19-MAR-1998.
XX 10-SEP-1997; 97WO-AU000591.
XX 27-FEB-1997; 97AU-0005374.
PR 10-SEP-1996; 96AU-0002262.
XX (AMRA-) AMRAD OPERATIONS PTY LTD.
XX Hilton DJ, Nicola NA, Simpson RJ, Zhang J;
XX WPI; 1998-207062/18.
DR N-PSDB; AAV22697.
XX
XX New isolated interleukin-13 binding protein - used to develop
PT products for therapy e.g. for allergic conditions such as asthma or
PT for diagnosis or detection
XX
XX Claim 7; Page 45-48; 69pp; English.
XX The IL-13 binding protein and related therapeutic molecules can be used
CC in the antagonism of at least one IL-13 activity. They can be used for
CC treating IL-13 mediated conditions such as certain allergic conditions
CC such as asthma or to inactivate locally administered IL-13 after IL-13
CC treatment. The products can also be used as diagnostic agents, e.g. for
CC detecting autoimmune diseases. The antibodies can also be used for
CC immunotherapy and may also be used as a diagnostic tool.
XX
XX Sequence 156 AA;
SQ
Query Match 37.5%; Score 3; DB 19; Length 156;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 sfxg 8
Db 101 sfxg 104
|||||
RESULT 12
AAB63508
ID AAB63508 standard; Protein; 174 AA.
XX
XX AAB63508;
XX 26-MAR-2001 (first entry)
XX Human gastric cancer associated antigen protein sequence SEQ ID NO:870.
DE Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW cancer associated antigen; cytostatic; cancer vaccine.
XX Homo sapiens.
OS
XX WO200073801-A2.
PN
XX 07-DEC-2000.
PD
XX 26-MAY-2000; 2000WO-US14749.
XX 28-MAY-1999; 99US-0136526.
PR 10-SEP-1999; 99US-0153454.
XX (LUDW-) LUDWIG INST CANCER RES.
PA Obata Y;
XX
PI WPI; 2001-025274/03.
DR
XX Nucleic acids encoding breast, gastric and prostate cancer associated

PT antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.
cancer -
XX
XX Example 1; Page 592-593; 799pp; English.
XX
XX AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
CC represent nucleotide sequences encoding human breast, gastric and
CC prostate cancer associated antigen precursors (CAAP) respectively.
CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC represent human breast, gastric and prostate CAAP protein sequence
CC respectively. CAAPs have cytostatic activity and can be used in the
CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic
CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC condition characterised by expression of an abnormal amount of a protein,
e.g. cancer.
XX
XX Sequence 174 AA;
SQ
Query Match 37.5%; Score 3; DB 22; Length 174;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 sfxg 8
Db 22 sfxg 25
|||||
RESULT 13
AAW28289
ID AAW28289 standard; Protein; 194 AA.
XX
XX AAW28289;
XX 14-SEP-1998 (first entry)
DT
XX Amino acid sequence of an enoyl-acyl carrier protein.
XX
XX Staphylococcus aureus protein; ribozyme; antisense sequence; control;
KW Staphylococcal gene; regulatory element; bacterial gene expression;
KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;
KW toxic shock syndrome.
XX
XX Staphylococcus aureus.
OS
XX Key Location/Qualifiers
FH Misc-difference 159 /note= "encoded by ANT"
FT Misc-difference 178 /note= "encoded by NTC"
FT Misc-difference 184 /note= "encoded by CNG"
FT
XX WO9730070-A1.
XX
XX 21-AUG-1997.
XX
XX 19-FEB-1997; 97WO-US02318.
PF
XX 20-FEB-1996; 96US-0011888.
PR
XX (SMIK) SMITHKLINE BEECHAM CORP.
PA
XX Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO;
PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;
XX WPI; 1997-424969/39.
DR N-PSDB; AAT84187.
XX
XX Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used
PT to isolate antimicrobial compounds, and in vaccines against S.
PT aureus infection

XX PS Claim 6; Page 586-587; 989pp; English.

CC The present sequence represents a Staphylococcus aureus protein, that,
CC based on homology with an E. coli protein, is believed to be an
CC enoyl-acyl carrier protein. The DNA sequence was isolated from a
CC library of clones of S. aureus WCUR 29 in Escherichia coli. The DNA
CC sequence can be used in the construction of ribozymes and antisense
CC sequences to control the expression of staphylococcal genes. The DNA
CC sequence is also useful as a source of regulatory elements for the
CC control of bacterial gene expression. The present protein may be used
CC to produce vaccines to enable a host to produce specific antibodies
CC with antibacterial action. These vaccines and antibodies would protect
CC a host against invasion by S. aureus, and conditions relating to
CC Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled
CC skin syndrome, and toxic shock syndrome.

XX SQ Sequence 194 AA;

Query Match 37.5%; Score 3; DB 18; Length 194;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 sfxg 8
Db 176 sfxg 179
||||

RESULT 14
AAB63791
ID AAB63791 standard; Protein; 217 AA.

XX AC AAB63791;

XX DT 26-MAR-2001 (first entry)

XX DE Human prostate cancer associated antigen protein sequence SEQ ID NO:1153.

XX DE Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW cancer associated antigen; cytostatic; cancer vaccine.

XX OS Homo sapiens.

XX PN WO200073801-A2.

XX PD 07-DEC-2000.

XX PF 26-MAY-2000; 2000WO-US14749.

XX PR 28-MAY-1999; 99US-0136526.

XX PR 10-SEP-1999; 99US-0153454.

XX PA (LUDW-) LUDWIG INST CANCER RES.

XX PI Obata Y;

XX DR WPI; 2001-025274/03.

XX Nucleic acids encoding breast, gastric and prostate cancer associated
PT antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.
PT cancer -

XX Example 1; Page 719; 799pp; English.

XX CC AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
CC represent nucleotide sequences encoding human breast, gastric and
CC prostate cancer associated antigen precursors (CAAP) respectively.
CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC represent human breast, gastric and prostate CAAP protein sequence
CC respectively. CAAPs have cytostatic activity and can be used in the
CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic

CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC condition characterised by expression of an abnormal amount of a protein,
CC e.g. cancer.

XX SQ Sequence 217 AA;

Query Match 37.5%; Score 3; DB 22; Length 217;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 sfxg 8
Db 214 sfxg 217
||||

RESULT 15
AAG72280
ID AAG72280 standard; Protein; 229 AA.

XX AC AAG72280;

XX DT 31-JUL-2001 (first entry)

XX DE Human olfactory receptor polypeptide, SEQ ID NO: 1961.

XX KW Human; olfactory receptor; OR; primary scent determination;
KW secondary scent determination; polypeptide library; odour receptor;
KW scent profile; scent fingerprint; scent representation.

XX OS Homo sapiens.

XX PN WO200127158-A2.

XX PD 19-APR-2001.

XX PF 06-OCT-2000; 2000WO-US27582.

XX PR 08-OCT-1999; 99US-0158615.

XX PR 24-FEB-2000; 2000US-0184809.

XX PA (DIGI-) DIGISCENTS.

XX PA (YEDA) YEDA RES & DEV CO LTD.

XX PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;

XX DR WPI; 2001-290713/30.

XX New polynucleotides which encode polypeptides involved in olfactory
PT sensation for identifying olfactory agonists and antagonists -

XX PS Claim 11; Page 1311-1312; 1857pp; English.

XX The present sequence is an olfactory receptor which is encoded by
CC one of a number of novel polynucleotides. The polynucleotides can be
CC used in screening for olfactory agonists and antagonists. The methods
CC allow for the determination of primary scents and the identification
CC of the odour receptors used to detect these primary scents. The methods
CC also enable determination of secondary scents and the identification of
CC combinations of odour receptors that are involved in detecting such
CC secondary scents. This enables the construction of a scent representation
CC (also called a scent fingerprint or scent profile), which may be used to
CC re-create and edit scents. Libraries of olfactory receptors are useful
CC for determining the interaction pattern of a composition with the
CC receptors, and can be used for determining differences in the olfactory
CC faculties of different individuals.

XX SQ Sequence 229 AA;

Query Match 37.5%; Score 3; DB 22; Length 229;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 sfxg 8
||||
Db 183 sfxg 186

RESULT 16
AAM41776
ID AAM41776 standard; Protein; 297 AA.
XX
XX AAM41776;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 6707.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.

OS Homo sapiens.

XX WO200153312-A1.

PN 26-JUL-2001.

PD 26-DEC-2000; 2000WO-US34263.

PF 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI: 2001-442253/47.

DR N-PSDB; AA160932.

XX Novel nucleic acids and polypeptides, useful for treating disorders

PT such as central nervous system injuries -

XX Example 2; SEQ ID NO 6707; 10078pp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AA38642-AA44213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies, and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX Sequence 297 AA;

Query Match 37.5%; Score 3; DB 22; Length 297;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 sfxg 8
||||
Db 256 sfxg 259

RESULT 17

AAG72750

ID AAG72750 standard; Protein; 312 AA.

XX AAG72750;

XX 31-JUL-2001 (first entry)

XX Human olfactory receptor data exploratorium sequence, SEQ ID NO: 2432.

XX Human; olfactory receptor; OR; primary scent determination;

KW secondary scent determination; polypeptide library; odour receptor;

KW scent profile; scent fingerprint; scent representation;

KW human olfactory receptor data exploratorium; HORDE.

XX Homo sapiens.

OS WO200127158-A2.

PN 19-APR-2001.

PD 06-OCT-2000; 2000WO-US27582.

PF 08-OCT-1999; 99US-0158615.

PR 24-FEB-2000; 2000US-0184809.

XX (DIGI-) DIGISCENTS.

PA (YEDA) YEDA RES & DEV CO LTD.

XX Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;

XX WPI: 2001-290713/30.

XX New polynucleotides which encode polypeptides involved in olfactory
PT sensation for identifying olfactory agonists and antagonists -

XX Example 6; Page 1652; 1857pp; English.

XX The present sequence is a polypeptide from the human olfactory receptor
CC data exploratorium (HORDE). It was used as a query sequence in a database
CC search of olfactory receptor (OR)-like sequences. The invention relates
CC to isolated polynucleotides encoding polypeptides involved in olfactory
CC sensation. The polynucleotides can be used in screening for olfactory
CC agonists and antagonists. The methods allow for the determination of
CC primary scents and the identification of the odour receptors used to
CC detect these primary scents. The methods also enable determination of
CC secondary scents and the identification of combinations of odour
CC receptors that are involved in detecting such secondary scents. This
CC enables the construction of a scent representation (also called a scent
CC fingerprint or scent profile), which may be used to re-create and edit
CC scents. Libraries of olfactory receptors are useful for determining the
CC interaction pattern of a composition with the receptors, and can be
CC used for determining differences in the olfactory faculties of different
CC individuals.

XX Sequence 312 AA;

Query Match 37.5%; Score 3; DB 22; Length 312;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 sfxg 8

Query Match 37.5%; Score 3; DB 21; Length 354;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 sfkg 8
Db 173 sfkg 176

RESULT 20

AAW48932
ID AAW48932 standard; protein: 500 AA.

XX AAW48932;

XX 08-OCT-1998 (first entry)

XX Human papillomavirus consensus sequence.

XX Human papillomavirus type 16; HPV16; virus-like particle; VLP;
KW HPV16D8202E; H16.04 antibody; H16.V5 antibody; identification;
KW HPV16 type; characterisation; HPV type 16 infection; HPV11; HPV6b.
XX Human papillomavirus.

OS Human papillomavirus.

XX Key Location/Qualifiers

FT Misc-difference 28 /note= "this residue is Lys in HPV11
FT and Thr in HPV6b"

FT Misc-difference 49 /note= "this residue is Tyr in HPV11
FT and Phe in HPV6b"

FT Misc-difference 53 /note= "this residue is Lys in HPV11
FT and Arg in HPV6b"

FT Misc-difference 54 /note= "this residue is Val in HPV11
FT and Ala in HPV6b"

FT Misc-difference 119 /note= "this residue is Leu in HPV11
FT and Phe in HPV6b"

FT Misc-difference 131 /note= "this residue is Gly in HPV11
FT and Ser in HPV6b"

FT Misc-difference 132 /note= "this residue is Tyr in HPV11"

FT Misc-difference 170 /note= "this residue is Thr in HPV11
FT and Lys in HPV6b"

FT Misc-difference 173 /note= "this residue is Ser in HPV11
FT and Thr in HPV6b"

FT Misc-difference 176 /note= "this residue is Ser in HPV11
FT and Pro in HPV6b"

FT Misc-difference 179 /note= "this residue is Asn in HPV11
FT and Arg in HPV6b"

FT Misc-difference 219 /note= "this residue is Leu in HPV11
FT and Ile in HPV6b"

FT Misc-difference 225 /note= "this residue is Val in HPV11
FT and Thr in HPV6b"

FT Misc-difference 246 /note= "this residue is Tyr in HPV11
FT and Phe in HPV6b"

FT Misc-difference 263 /note= "this residue is Thr in HPV11
FT and Glu in HPV6b"

FT Misc-difference 271

FT /note= "this residue is Asp in HPV11
FT and Thr in HPV6b"
FT Misc-difference 273 /note=
FT "this residue is Leu in HPV11
FT and Ile in HPV6b"
FT Misc-difference 274 /note=
FT "this residue is Val in HPV11
FT and Ile in HPV6b"
FT Misc-difference 277 /note=
FT "this residue is Gly in HPV11
FT and Ser in HPV6b"
FT Misc-difference 278 /note=
FT "this residue is Asn in HPV11
FT and Gly in HPV6b"
FT Misc-difference 281 /note=
FT "this residue is Ser in HPV11
FT and Thr in HPV6b"
FT Misc-difference 284 /note=
FT "this residue is Ala in HPV11
FT and Gly in HPV6b"
FT Misc-difference 290 /note=
FT "this residue is His in HPV11
FT and Asn in HPV6b"
FT Misc-difference 325 /note=
FT "this residue is His in HPV11
FT and Gln in HPV6b"
FT Misc-difference 346 /note=
FT "this residue is Ser in HPV11
FT and Lys in HPV6b"
FT Misc-difference 347 /note=
FT "this residue is Lys in HPV11
FT and Thr in HPV6b"
FT Misc-difference 349 /note=
FT "this residue is Ala in HPV11
FT and Ser in HPV6b"
FT Misc-difference 366 /note=
FT "this residue is Phe in HPV11
FT and Tyr in HPV6b"
FT Misc-difference 434 /note=
FT "this residue is Gln in HPV11
FT and Pro in HPV6b"
FT Misc-difference 439 /note=
FT "this residue is Asp in HPV11
FT and Asn in HPV6b"
FT Misc-difference 440 /note=
FT "this residue is Met in HPV11
FT and Leu in HPV6b"
FT Misc-difference 458 /note=
FT "this residue is Phe in HPV11
FT and Tyr in HPV6b"
FT Misc-difference 474 /note=
FT "this residue is Thr in HPV11
FT and Ser in HPV6b"
FT Misc-difference 476 /note=
FT "this residue is Ala in HPV11
FT and Ile in HPV6b"
FT Misc-difference 480 /note=
FT "this residue is Ile in HPV11
FT and Val in HPV6b"
FT Misc-difference 488 /note=
FT "this residue is Pro in HPV11
FT and Ala in HPV6b"
FT Misc-difference 490 /note=
FT "this residue is Thr in HPV11
FT and Ala in HPV6b"
FT Misc-difference 497 /note=
FT "this residue is Thr in HPV11
FT and Ala in HPV6b"

XX WO9825646-A1.

XX 18-JUN-1998.

XX

PF 05-DEC-1997; 97WO-US22023.
XX
PR 09-DEC-1996; 96US-0032633.
XX
PA (MERI) MERCK & CO INC.
XX
XX Ludmerer S;
XX
XX WPI; 1998-348266/30.
XX
XX Human papillomavirus type 16 D202E virus-like particles - used for
PT characterising HPV16 type infections
XX
XX Disclosure; Fig 1; 37pp; English.
XX
XX The present sequence represents a consensus sequence between human
CC papillomavirus type 11 (HPV11) and HPV6b. The specification describes a
CC synthetic HPV16 virus-like particle (VLP) which contains a Glu
CC substitution for the naturally occurring Asp at position 202. The protein
CC is designated HPV16Dd202E. The novel VLP HPV:D202E binds H16.U4 antibody
CC but not H16.V5 antibody and thus allows specific identification of HPV16
CC types. The new VLPs are used in the characterisation of HPV type 16
CC infections.
XX
SQ Sequence 500 AA;

Query Match 37.5%; Score 3; DB 19; Length 500;
Best Local Similarity 100.0%; Pred. No. 2.4e+03; Indels 0; Gaps 0;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 kxxsf 6
Db 438 kxxsf 442

RESULT 21
AAB48242
ID AAB48242 standard; Protein; 552 AA.
XX
AC AAB48242;
XX
XX 02-APR-2001 (first entry)
XX
XX Amino acid sequence of bZIP2 ORF2 protein.
XX
XX
XX Transcription factor; seed storage protein; lectin; oil-body protein;
KW Pv-Seed factor-1; ROM1; Vicilin-box binding protein-1; ROM2; 7S-globulin;
KW Phaseolin; PHA-L; bean; nuclear protein; promoter; ORF; bZIP;
KW basic leucine zipper.
XX
XX Phaseolus vulgaris.
XX
XX Key Location/Qualifiers
FH Misc-difference 1..552
FT /note= "Xaa are residues encoded by internal stop codons"
FT
XX
XX US6160202-A.
XX
XX 12-DEC-2000.
XX
XX 06-FEB-1997; 97US-0796899.
XX
XX 07-OCT-1994; 94US-0319544.
XX
XX (UYMA-) UNIV MARYLAND BALTIMORE COUNTY.
XX
XX Chern M, Bustos MM;
XX
XX WPI; 2001-079619/09.
XX
XX N-PSDB; AAC84565.
XX
XX Novel transcription factor gene which encodes transcription factor

PT protein that targets promoters of genes encoding seed storage proteins
PT are useful for modulating seed storage protein expression in dicot seed
PT crops -
XX
XX Disclosure; Columns 35-38; 67pp; English.
XX
XX The invention relates to an isolated transcription factor gene which is
CC expressed in a recombinant maturing dicot seed and which encodes a
CC transcription factor protein which targets a promoter of a gene encoding
CC seed storage proteins, lectins or oil-body proteins. The transcription
CC factors isolated are Pv-Seed factor-1 (ROM1) and Vicilin-box binding
CC protein-1 (ROM2). These factors bind to 7S-globulin (b-phaseolin) or
CC lectin (PHA-L) promoters. The transcription factor gene is useful for
CC enhancing or reducing expression of seed storage protein, lectin or
CC oil-protein genes in dicot seed crops. The present sequence represents
CC the amino acid sequence of bZIP2 (basic leucine zipper) ORF2 protein.
XX
SQ Sequence 552 AA;

Query Match 37.5%; Score 3; DB 22; Length 552;
Best Local Similarity 100.0%; Pred. No. 2.5e+03; Indels 0; Gaps 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 sfxg 8
Db 278 sfxg 281

RESULT 22
AAR85443
ID AAR85443 standard; Protein; 908 AA.
XX
XX AAR85443;
AC
XX
XX 19-FEB-1996 (first entry)
XX
XX Rat brevicin core protein.
XX
XX Brevican; chondroitin sulphate proteoglycan; glial cell; axon;
KW neurofibromatosis; gliosis.
XX
XX Rattus sp.
XX
XX Key Location/Qualifiers
FH Misc-difference 41..50
FT /note= "unidentified amino acids"
FT Misc-difference 72
FT /note= "unidentified amino acid"
FT Misc-difference 74..75
FT /note= "unidentified amino acids"
FT Misc-difference 93..109
FT /note= "unidentified amino acids"
FT Misc-difference 115
FT /note= "unidentified amino acid"
FT Misc-difference 117..120
FT /note= "unidentified amino acids"
FT Misc-difference 138
FT /note= "unidentified amino acid"
FT Misc-difference 150
FT /note= "unidentified amino acid"
FT Misc-difference 155
FT /note= "unidentified amino acid"
FT Misc-difference 162..166
FT /note= "unidentified amino acids"
FT Misc-difference 173
FT /note= "unidentified amino acid"
FT Misc-difference 201..222
FT /note= "unidentified amino acids"
FT Misc-difference 290
FT /note= "unidentified amino acid"
FT Misc-difference 300
FT /note= "unidentified amino acid"

FT Misc-difference 324 /note= "unidentified amino acid"
FT Misc-difference 395 /note= "unidentified amino acid"
FT Misc-difference 413 /note= "unidentified amino acid"
FT Misc-difference 439..440 /note= "unidentified amino acids"
FT Misc-difference 447..490 /note= "unidentified amino acids"
FT Misc-difference 524..530 /note= "unidentified amino acids"
FT Misc-difference 561..581 /note= "unidentified amino acids"
FT Misc-difference 603 /note= "unidentified amino acids"
FT Misc-difference 638..641 /note= "unidentified amino acid"
FT Misc-difference 718..720 /note= "unidentified amino acids"
FT Misc-difference 751..800 /note= "unidentified amino acids"
FT Misc-difference 835..836 /note= "unidentified amino acids"
FT Misc-difference 866..867 /note= "unidentified amino acids"
FT Misc-difference 897..898 /note= "unidentified amino acids"
XX WO9526201-A1.
XX
XX
XX 05-OCT-1995.
XX 27-MAR-1995; 95WO-US03747.
XX 28-MAR-1994; 94US-0219642.
XX (LJOL-) LA JOLLA CANCER RES FOUND.
XX Shimonaka M, Watanabe K, Yamada H, Yamaguchi Y;
XX WPI; 1995-351200/45.
XX Mammalian brevican protein - directs/inhibits axonal growth, used in
XX treatment of neuro-fibromatosis and in detection of gliosis
XX Claim 6; Page 45-47; 73pp; English.
XX Brevican was isolated from the proteoglycan soluble fraction of
XX rat brain. The fraction containing the core protein was
XX purified and the sequences of the N-terminus and internal tryptic
XX peptides were detd., giving the sequence shown in AAR85443.
XX Brevican is used to modulate axonal growth and to raise antibodies
XX useful for detecting gliosis.
SQ Sequence 908 AA;

Query Match 37.5%; Score 3; DB 16; Length 908;
Best Local Similarity 100.0%; Pred. No. 3.5e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sfxg 8
DB 171 sfxg 174

RESULT 23
AAV05839
ID AAV05839 standard; Protein; 1186 AA.
XX
XX
AC AAV05839;
XX

DT 02-AUG-1999 (first entry)
XX
DE Banana ripening fruit Gluc. translated polypeptide.
XX
KW Banana; fruit ripening; glucanase; differential expression;
KW fruit development; transgenic plant.
XX
OS Musa acuminata.
XX
XX
FH Key Location/Qualifiers
FT Misc-difference 1..1186 /note= "the X residues in this sequence correspond
FT to in-frame stop codons in reading frame 1
FT of Gluc. DNA"
XX
XX WO9915668-A2.
XX
XX 01-APR-1999.
XX
XX 23-SEP-1998; 98WO-US03343.
XX 25-SEP-1997; 97US-0060062.
XX (BOYC-) BOYCE THOMPSON INST PLANT RES.
XX
XX Clendennen S, May G;
XX WPI; 1999-244425/20.
XX N-PSDB; AAX25612.
XX
XX New isolated banana DNA molecules
XX
XX Disclosure; Fig 15A-E; 143pp; English.
XX
XX The present sequence represents the protein encoded by reading
XX frame 1 of a DNA sequence termed Gluc. (see AAX25612) of the
XX invention. Amino acid sequences deduced from reading frames 2 and
XX 3 are given in AAY05840 and AAY05841, respectively. The identity of
XX GLUC. and the coding region of GLUC. DNA are not indicated. The
XX invention provides isolated DNA molecules which are differentially
XX expressed during banana fruit development, and the protein products
XX of these genes. The DNA is selected from a group comprising starch
XX synthase, chitinase, endochitinase, beta-1,3-glucanase,
XX thaumatin-like protein, ascorbate peroxidase, metallothionein,
XX lectin and senescence-related protein. The regulatory elements of
XX the genes can be used to produce chimeric genes for transformation
XX of plants to provide controlled expression of heterologous DNA
XX during fruit development, or in response to exogenous developmental
XX signals, such as ethylene signals. The heterologous protein, e.g.
XX a therapeutic protein, can be isolated from the fruit or consumed
XX directly in the transformed fruit.
XX
SQ Sequence 1186 AA;

Query Match 37.5%; Score 3; DB 20; Length 1186;
Best Local Similarity 100.0%; Pred. No. 4.1e+03;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 kxxsf 6
DB 117 kxxsf 121

RESULT 24
AAV05842
ID AAY05842 standard; Protein; 2466 AA.
XX
XX AAY05842;
XX
XX 02-AUG-1999 (first entry)
XX
XX Banana ripening fruit ENDO. translated polypeptide.
DE

XX Banana; fruit ripening; glucanase; differential expression;
KW fruit development; transgenic plant.
XX Musa acuminata.
OS
XX
XX
FH Key Location/Qualifiers
FT Misc-difference 1..2466
FT /note= "the X residues in this sequence correspond
FT to in-frame stop codons or degenerate
FT codons in reading frame 1 of ENDO. DNA"
XX
XX WO9915668-A2.
FN
PN
XX
XX 01-APR-1999.
XX
XX 23-SEP-1998; 98WO-US03343.
PF
XX 25-SEP-1997; 97US-0060062.
XX
XX (BOVC-) BOYCE THOMPSON INST PLANT RES.
PA
XX Clendennen S, May G;
PI
XX WPI; 1999-244425/20.
DR
DR N-PSDB; AAX25613.
XX
XX New isolated banana DNA molecules
PT
XX Claim 14; Fig 16A-J; 143pp; English.
PS
XX The present sequence represents the protein encoded by reading
CC frame 1 of a DNA sequence termed ENDO. (see AAX25613) of the
CC invention. Amino acid sequences deduced from reading frames 2 and
CC 3 are given in AAY05843 and AAY05844, respectively. The identity of
CC ENDO. and the coding region of ENDO. DNA are not indicated. The
CC invention provides isolated DNA molecules which are differentially
CC expressed during banana fruit development, and the protein products
CC of these genes. The DNA is selected from a group comprising starch
CC synthase, chitinase, endochitinase, beta-1,3-glucanase,
CC thaumatin-like protein, ascorbate peroxidase, metallothionein,
CC lectin and senescence-related protein. The regulatory elements of
CC the genes can be used to produce chimeric genes for transformation
CC of plants to provide controlled expression of heterologous DNA
CC during fruit development, or in response to exogenous developmental
CC signals, such as ethylene signals. The heterologous protein, e.g.
CC a therapeutic protein, can be isolated from the fruit or consumed
CC directly in the transformed fruit.
XX
SQ Sequence 2466 AA;

Query Match 37.5%; Score 3; DB 20; Length 2466;
Best Local Similarity 100.0%; Pred. No. 6.4e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 sfxg 8
|||
Db 1912 sfxg 1915

RESULT 25
AAR42556
ID AAR42556 standard; peptide; 3 AA.
XX
AC AAR42556;
XX
XX 07-DEC-1994 (first entry)
XX ACE inhibitor SP3.
XX
XX Angiotensin converting enzyme; ACE; inhibitor; SP5; SP4; SP3;
KW hypertension; blood pressure.

XX Synthetic.
OS
XX JP06065288-A.
PN
XX
XX 08-MAR-1994.
PD
XX 19-AUG-1992; 92JP-0220270.
PF
XX 19-AUG-1992; 92JP-0220270.
PR
XX (APIA-) API KK.
PA
XX WPI; 1994-115194/14.
DR
XX
XX New tri-, tetra- and penta-peptide(s), e.g. Trp-Lys-Tyr - are ACE
PT inhibitors useful for treatment or prophylaxis of hypertension
PT
XX Claim 1; Page 2; 5pp; Japanese.
PS
XX Peptides SP3, SP4 and SP5 have ACE inhibiting activity. They can be
CC prepd. easily and in high yield. They are useful for treatment or
CC prophylaxis of hypertension.
CC
XX Sequence 3 AA;
SQ

Query Match 25.0%; Score 2; DB 15; Length 3;
Best Local Similarity 100.0%; Pred. No. 4.3e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 wk 2
||
Db 1 wk 2

Search completed: January 14, 2002, 07:56:30
Job time: 725 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:57:30 ; Search time 41.59 Seconds
(without alignments)
4.329 Million cell updates/sec

Title: 09-185908-1c
Perfect score: 8
Sequence: 1 wxxxxfxg 8

Scoring table:
OLIGO
Gapop 60.0 , Capext 60.0
Searched: 212252 seqs, 22503292 residues

Word size : 0
Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 50 summaries

- Database : Issued Patents AA:*
- 1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
 - 2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
 - 3: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
 - 4: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
 - 5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*
 - 6: /cgn2_6/ptodata/2/1aa/Backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37.5	7	2	US-08-764-157-8	Sequence 8, Appli
2	37.5	31	3	US-09-295-186-8	Sequence 8, Appli
3	37.5	501	2	US-08-987-519-3	Sequence 3, Appli
4	37.5	552	4	US-08-796-899-28	Sequence 28, Appli
5	37.5	908	5	PCT-US95-03747-3	Sequence 3, Appli
6	25.0	2	3	US-08-889-841B-3	Sequence 6, Appli
7	25.0	2	3	US-08-889-841B-6	Sequence 13, Appli
8	25.0	2	3	US-08-889-841B-13	Sequence 17, Appli
9	25.0	2	3	US-08-889-841B-17	Sequence 20, Appli
10	25.0	2	3	US-08-889-841B-20	Sequence 34, Appli
11	25.0	2	3	US-08-889-841B-34	Sequence 37, Appli
12	25.0	2	3	US-08-889-841B-37	Sequence 40, Appli
13	25.0	2	3	US-08-889-841B-40	Sequence 405, App
14	25.0	2	3	US-09-461-697-405	Sequence 14, Appli
15	25.0	2	3	US-07-657-769B-14	Sequence 11, Appli
16	25.0	2	3	US-07-714-540-11	Sequence 12, Appli
17	25.0	2	3	US-07-714-540-12	Sequence 38, Appli
18	25.0	2	3	US-07-828-450-38	Sequence 7, Appli
19	25.0	2	3	US-08-079-445-3	Sequence 11, Appli
20	25.0	2	3	US-07-840-077A-7	Sequence 12, Appli
21	25.0	2	3	US-08-127-904-11	Sequence 12, Appli
22	25.0	2	3	US-08-127-904-12	Sequence 12, Appli
23	25.0	2	3	US-08-405-933-14	Sequence 12, Appli
24	25.0	2	3	US-08-461-611-12	Sequence 109, App
25	25.0	2	3	US-07-789-184-109	Sequence 117, App
26	25.0	2	3	US-07-789-184-117	Sequence 7, Appli
27	25.0	2	3	US-08-454-950-7	

28	2	25.0	4	1	US-08-434-761-3	Sequence 3, Appli
29	2	25.0	4	1	US-08-338-890B-1	Sequence 1, Appli
30	2	25.0	4	1	US-08-454-949-7	Sequence 7, Appli
31	2	25.0	4	1	US-08-475-263-109	Sequence 109, App
32	2	25.0	4	1	US-08-475-263-117	Sequence 117, App
33	2	25.0	4	1	US-08-485-886-109	Sequence 109, App
34	2	25.0	4	1	US-08-485-886-117	Sequence 117, App
35	2	25.0	4	2	US-08-441-871-62	Sequence 62, Appli
36	2	25.0	4	2	US-08-441-871-72	Sequence 72, Appli
37	2	25.0	4	2	US-08-477-362-109	Sequence 109, App
38	2	25.0	4	2	US-08-477-362-117	Sequence 117, App
39	2	25.0	4	2	US-08-477-134-109	Sequence 109, App
40	2	25.0	4	2	US-08-477-134-117	Sequence 117, App
41	2	25.0	4	2	US-08-671-487A-8	Sequence 8, Appli
42	2	25.0	4	2	US-08-350-260A-408	Sequence 408, App
43	2	25.0	4	2	US-08-340-208B-1	Sequence 1, Appli
44	2	25.0	4	2	US-09-195-049-4	Sequence 4, Appli
45	2	25.0	4	2	US-08-651-179B-2	Sequence 2, Appli
46	2	25.0	4	2	US-08-651-179B-3	Sequence 3, Appli
47	2	25.0	4	3	US-08-467-580-82	Sequence 82, Appli
48	2	25.0	4	3	US-08-997-263-1	Sequence 1, Appli
49	2	25.0	4	3	US-08-473-489A-109	Sequence 109, App
50	2	25.0	4	3	US-08-473-489A-117	Sequence 117, App

ALIGNMENTS

RESULT 1
US-08-764-157-8
; Sequence 8, Application US/08764157
; Patent No. 5830863
; GENERAL INFORMATION:
; APPLICANT: Buck, Stephen H
; APPLICANT: Harteson, Scott L
; APPLICANT: Kristiansky, John L
; APPLICANT: Hassman III, Chester F
; APPLICANT: McCarthy, James Ray
; TITLE OF INVENTION: Neurokinin A Antagonists
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marion Merrell Dow Inc.
; STREET: 2110 East Galbraith Rd.
; CITY: Cincinnati P. O. Box 156300
; STATE: Ohio
; COUNTRY: USA
; ZIP: 45215-6300
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,157
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 07/686,593
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Nesbitt, Stephen L
; REGISTRATION NUMBER: 28981
; REFERENCE/DOCKET NUMBER: M01352C US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (513) 948-7965
; TELEFAX: (513) 948-7961
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:

```
; NAME/KEY: Modified-site
; LOCATION: 4
; OTHER INFORMATION: /note="Xaa is a Valine analog
; OTHER INFORMATION: having a 1-methylene group, in place of a
; OTHER INFORMATION: 1-carbonyl group, bonded to the alpha nitrogen"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 4
; OTHER INFORMATION: /note="(cont'd) of the subsequent
; OTHER INFORMATION: amino acid"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 7
; OTHER INFORMATION: /note="Xaa is Leucin-1-amide
; OTHER INFORMATION: (Leu-NH2)"
; US-08-764-157-8

Query Match          37.5%; Score 3; DB 2; Length 7;
Best Local Similarity 100.0%; Pred No. 1.6e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sfxg 8
    ||||
Db 2 SFXG 5

RESULT 2
US-09-295-186-8
; Sequence 8, Application US/09295186B
; Patent No. 6127137
; GENERAL INFORMATION:
; APPLICANT: Hasida, Miyoko
; APPLICANT: Tsutsumi, No. 6127137iko
; APPLICANT: Halkier, Torben
; APPLICANT: Stringer, Mary Ann
; TITLE OF INVENTION: An Acidic Phospholipase, Production, and
; TITLE OF INVENTION: Methods of Using Thereof (As Amended)
; FILE REFERENCE: 4953, 204-US
; CURRENT APPLICATION NUMBER: US/09/295,186B
; CURRENT FILING DATE: 1999-04-20
; PRIOR APPLICATION NUMBER: 1215/96
; PRIOR FILING DATE: 1996-10-31
; PRIOR APPLICATION NUMBER: PCT/DK97/00490
; PRIOR FILING DATE: 1997-10-30
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Hypochozyna sp. CBS 648.91
; FEATURE:
; OTHER INFORMATION: Xaa at position 5 is any amino acid
; US-09-295-186-8

Query Match          37.5%; Score 3; DB 3; Length 31;
Best Local Similarity 100.0%; Pred No. 69;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sfxg 8
    ||||
Db 3 SFXG 6

RESULT 3
US-08-987-519-3
; Sequence 3, Application US/08987519
; Patent No. 5952216
; GENERAL INFORMATION:
; APPLICANT: Ludmerer, Steven
; TITLE OF INVENTION: Synthetic HPV16 Virus-Like Particles
; FILE REFERENCE: 19853
```

```
; CURRENT APPLICATION NUMBER: US/08/987,519
; CURRENT FILING DATE: 1997-12-09
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: SIMILAR
; LOCATION: (0)...(0)
; OTHER INFORMATION: Consensus Sequence
; OTHER INFORMATION: At 28 X = K or T
; OTHER INFORMATION: At 49 X = Y or F
; OTHER INFORMATION: At 53 X = K or R
; OTHER INFORMATION: At 54 X = V or A
; OTHER INFORMATION: At 119 X = L or F
; OTHER INFORMATION: At 131 X = G or S
; OTHER INFORMATION: At 132 X = Y or N
; OTHER INFORMATION: At 170 X = T or K
; OTHER INFORMATION: At 173 X = S or T
; OTHER INFORMATION: At 176 X = S or P
; OTHER INFORMATION: At 179 X = N or A
; OTHER INFORMATION: At 219 X = L or I
; OTHER INFORMATION: At 225 X = V or T
; OTHER INFORMATION: At 246 X = Y or F
; OTHER INFORMATION: At 263 X = T or E
; OTHER INFORMATION: At 271 X = D or T
; OTHER INFORMATION: At 273 X = L or I
; OTHER INFORMATION: At 274 X = V or I
; OTHER INFORMATION: At 277 X = G or S
; OTHER INFORMATION: At 278 X = N or G
; OTHER INFORMATION: At 281 X = S or T
; OTHER INFORMATION: At 284 X = A or G
; OTHER INFORMATION: At 290 X = H or N
; OTHER INFORMATION: At 325 X = H or Q
; OTHER INFORMATION: At 346 X = S or T
; OTHER INFORMATION: At 347 X = K or T
; OTHER INFORMATION: At 348 X = A or S
; OTHER INFORMATION: At 366 X = F or Y
; OTHER INFORMATION: At 434 X = Q or P
; OTHER INFORMATION: At 439 X = D or N
; OTHER INFORMATION: At 440 X = M or L
; OTHER INFORMATION: At 448 X = F or Y
; OTHER INFORMATION: At 458 X = F or L
; OTHER INFORMATION: At 474 X = T or S
; OTHER INFORMATION: At 476 X = A or I
; OTHER INFORMATION: At 480 X = I or V
; OTHER INFORMATION: At 488 X = P or A
; OTHER INFORMATION: At 490 X = T or A
; OTHER INFORMATION: At 497 X = T or A
; OTHER INFORMATION: At 501 X = K or R
; US-08-987-519-3

Query Match          37.5%; Score 3; DB 2; Length 501;
Best Local Similarity 100.0%; Pred No. 6.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 kxxsf 6
    |||||
Db 438 KXXSF 442

RESULT 4
US-08-796-899-28
; Sequence 28, Application US/08796899
; Patent No. 6160202
; GENERAL INFORMATION:
; APPLICANT: BUSTOS, Mauricio M
; APPLICANT: CHERN, Maw-Sheng
; TITLE OF INVENTION: MODIFICATION OF SEED CROPS WITH
; TITLE OF INVENTION: TRANSCRIPTION FACTORS
; NUMBER OF SEQUENCES: 32
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:58:37 ; Search time 45.4 Seconds
(without alignments)
13.423 Million cell updates/sec

Title: 09-185908-1E
Perfect score: 8
Sequence: 1 wrxxxysxg 8

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 50 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2	25.0	4	2 A37832	phenol 2-monooxyge
2	2	25.0	2	2 A10469	dnazx-like protein
3	2	25.0	7	1 NYPG7	hypothalamic hepta
4	2	25.0	7	2 A60139	fatty-acid synthas
5	2	25.0	7	2 S33245	neuromodulatory pe
6	2	25.0	7	4 I56695	hypothetical L2 pr
7	2	25.0	8	2 S11078	glucose-6-phosphat
8	2	25.0	8	2 B27867	homeotic protein U
9	2	25.0	9	2 B57444	neuropeptide Grb-A
10	2	25.0	9	2 C57444	neuropeptide Grb-A
11	2	25.0	9	2 PT0238	Ig heavy chain CRD
12	2	25.0	9	2 PT0324	Ig heavy chain CRD
13	2	25.0	9	2 I46016	cytochrome 4 - bo
14	2	25.0	9	2 A28924	fructose-bisphosph
15	2	25.0	9	2 S36850	Ig heavy chain V r
16	2	25.0	9	2 G41946	T-cell receptor ga
17	2	25.0	10	1 RHPCG	gonadoliblerin - pi
18	2	25.0	10	1 RHSHG	gonadoliblerin - sh
19	2	25.0	10	1 RHAQ1	gonadoliblerin I -
20	2	25.0	10	2 A21114	gonadoliblerin - ch
21	2	25.0	10	2 PH0948	T-cell receptor be
22	2	25.0	11	2 A40693	transgelin - sheep
23	2	25.0	11	2 A34662	Achactina cardio-ex
24	2	25.0	11	2 B41946	T-cell receptor ga
25	2	25.0	11	2 I60434	68kDa neurofilamen
26	2	25.0	11	2 PH0941	T-cell receptor be
27	2	25.0	11	2 PH0938	T-cell receptor be
28	2	25.0	11	2 PH0914	T-cell receptor be
29	2	25.0	12	1 A53709	alpha-conotoxin Im

30	2	25.0	12	2 A28856	fructose-bisphosph
31	2	25.0	12	2 S26546	T-cell receptor be
32	2	25.0	12	2 S26556	T-cell receptor be
33	2	25.0	12	2 S26552	T-cell receptor be
34	2	25.0	12	2 S25056	Ig heavy chain - m
35	2	25.0	12	2 A49261	coagulation factor
36	2	25.0	12	2 S47391	T-cell antigen rec
37	2	25.0	12	2 S47395	T-cell antigen rec
38	2	25.0	12	2 PH1611	Ig H chain V-D-J r
39	2	25.0	12	2 PH1461	T-cell receptor be
40	2	25.0	12	2 PH1459	T-cell receptor be
41	2	25.0	12	2 PH0936	T-cell receptor be
42	2	25.0	13	1 MTCMAD	melanotropin alpha
43	2	25.0	13	1 MTHOAD	melanotropin alpha
44	2	25.0	13	2 PC1149	equinatoxin 1A - s
45	2	25.0	13	2 A33660	osteoclast functio
46	2	25.0	13	2 PR0263	Ig heavy chain CRD
47	2	25.0	13	2 S47358	T-cell antigen rec
48	2	25.0	13	2 S47359	T-cell antigen rec
49	2	25.0	13	2 S47362	T-cell antigen rec
50	2	25.0	13	2 S47365	T-cell antigen rec

ALIGNMENTS

RESULT 1
A37832
phenol 2-monooxygenase (EC 1.14.13.7) chain P5 - Pseudomonas sp. (strain CP600) (frag
C:Species: Pseudomonas sp.
C:Date: 14-Jun-1991 #sequence_revision 14-Jun-1991 #text_change 23-Jun-1993
C:Accession: A37832
R:Powlowski, J.; Shingler, V.
J. Bacteriol. 172, 6834-6840, 1990
A:Title: In vitro analysis of polypeptide requirements of multicomponent phenol hydro
A:Reference number: A37832; MUID:91072231
A:Accession: A37832
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-4 <POM>
C:Keywords: oxidoreductase

Query Match 25.0%; Score 2; DB 2; Length 4;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
Db 1 sy 2

RESULT 2
I40469
dnazx-like protein - Bacillus subtilis (fragment)
C:Species: Bacillus subtilis
C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 21-Jul-2000
C:Accession: I40469
R:Struck, J.C.; Hartmann, R.K.; Tosechka, H.Y.; Erdmann, V.A.
Mol. Gen. Genet. 215, 478-482, 1989
A:Title: Transcription and processing of Bacillus subtilis small cytoplasmic RNA.
A:Reference number: I40469; MUID:89218958
A:Accession: I40469
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-5 <RES>
A:Cross-references: EMBL:X14796; NID:940130; PIDN:CAA32902.1; PID:94376204
C:Genetics:
A:Start codon: GTG

Query Match 25.0%; Score 2; DB 2; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;

Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
||

Db 2 sy 3

RESULT 3

NYPG7

hypothalamic heptapeptide - pig

C:Species: Sus scrofa domestica (domestic pig)

C>Date: 01-Sep-1981 #sequence_revision 01-Sep-1981 #text_change 23-Aug-1996

C:Accession: A01417

R:Chang, R.C.C.; Huang, W.Y.; Arimura, A.; Redding, T.W.; Coy, D.H.; Saffran, M.; Kong,

Horm. Metab. Res. 13, 228-232, 1981

A:Title: Isolation, structure and synthesis of a heptapeptide with in vitro ACTH-release

A:Reference number: A01417; MUID:81213980

A:Accession: A01417

A:Molecule type: protein

A:Residues: 1-7 <CHA>

C:Superfamily: hypothalamic heptapeptide

C:Keywords: hypothalamus

Query Match 25.0%; Score 2; DB 1; Length 7;

Best Local Similarity 100.0%; Pred. No. 2.2e+05;

Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
||

Db 5 sy 6

RESULT 4

A60139

fatty-acid synthase (EC 2.3.1.85) - rabbit (fragment)

C:Species: Oryctolagus cuniculus (domestic rabbit)

C>Date: 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change 26-May-2000

C:Accession: A60139

R:Hardie, D.G.; Dewar, K.B.; Altken, A.; McCarthy, A.D.

Biochim. Biophys. Acta 828, 380-382, 1985

A:Title: Amino acid sequence around the reactive serine residue of the thioesterase domain

A:Reference number: A60139; MUID:85175165

A:Accession: A60139

A:Molecule type: protein

A:Residues: 1-7 <HNR>

C:Superfamily: rat fatty-acid synthase; 3-oxoacyl-[acyl-carrier-protein] synthase I hom

C:Keywords: acyltransferase; short-chain alcohol dehydrogenase homology; [acyl-carrier-protein] S-

F:5/Active site: Ser (of oleoyl-[acyl-carrier-protein] hydrolyase) #status experimental

Query Match 25.0%; Score 2; DB 2; Length 7;

Best Local Similarity 100.0%; Pred. No. 2.2e+05;

Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
||

Db 5 sy 6

RESULT 5

S33245

neuromodulatory peptide Wkamide-2 - giant African snail

C:Species: Achatina fulica (giant African snail)

C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Jul-1997

C:Accession: S33245

R:Minakata, H.; Ikeda, T.; Muneoka, Y.; Kobayashi, M.; Nomoto, K.

FEBS Lett. 323, 104-108, 1993

A:Title: Wkamide-1, -2 and -3: novel neuromodulatory peptides isolated from ganglia of t

A:Reference number: S33244; MUID:93265912

A:Accession: S33245

A:Status: preliminary

A:Molecule type: protein
A:Residues: 1-7 <MIN>

Query Match 25.0%; Score 2; DB 2; Length 7;

Best Local Similarity 100.0%; Pred. No. 2.2e+05;

Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wr 2
||

Db 1 wr 2

RESULT 6

I56695

hypothetical L2 protein (mistranslated) - human papillomavirus type 16 (fragment)

C:Species: human papillomavirus type 16

C>Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 18-Feb-2000

C:Accession: I56695

R:Schneider-Maunoury, S.; Croissant, O.; Orth, G.

J. Virol. 61, 3295-3298, 1987

A:Title: Integration of human papillomavirus type 16 DNA sequences: a possible early

A:Reference number: I56695; MUID:87311896

A:Accession: I56695

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-7 <SCH>

A:Cross-references: GB:M30709; MID:q190253; PIDN:AAA65995.1; PID:q553616

C:Comment: This is the hypothetical translation of a viral sequence integrated into t

C:Comment: It is translated in an incorrect, -1, reading frame of the L2 protein.

Query Match 25.0%; Score 2; DB 4; Length 7;

Best Local Similarity 100.0%; Pred. No. 2.2e+05;

Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
||

Db 2 sy 3

RESULT 7

S11078

glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) - yeast (Pichia jadinii) (fragment)

C:Species: Pichia jadinii, Candida utilis

C>Date: 30-Jun-1991 #sequence_revision 30-Sep-1991 #text_change 05-Aug-1994

C:Accession: S11078

R:Eggestad, B.; Estonius, M.; Danielsson, O.; Persson, B.; Cederlund, E.; Kaiser, R.;

FEBS Lett. 269, 194-196, 1990

A:Title: Fast atom bombardment mass spectrometry and chemical analysis in determinati

A:Reference number: S11074; MUID:90353571

A:Accession: S11078

A:Molecule type: protein

A:Residues: 1-8 <EGE>

A:Note: the source is designated as Pichia jadinii

C:Keywords: acetylated amino end; oxidoreductase; pentose phosphate pathway

F:1/Modified site: acetylated amino end (Ser) #status experimental

Query Match 25.0%; Score 2; DB 2; Length 8;

Best Local Similarity 100.0%; Pred. No. 2.2e+05;

Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
||

Db 1 sy 2

RESULT 8

B27867

homeotic protein Ultrabithorax - fruit fly (Drosophila melanogaster) (fragment)

C:Species: Drosophila melanogaster

C>Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 23-Feb-1997

C:Accession: B27867
R:Saari, G.; Blenz, M.
EMBO J. 6, 1775-1779, 1987
A:Title: The structure of the ultrabithorax promoter of *Drosophila melanogaster*.
A:Reference number: A91072
A:Accession: B27867
A:Molecule type: mRNA
A:Residues: 1-8 <SAS>
C:Genetics:
A:Gene: FlyBase:Ubx
A:Cross-references: FlyBase:FBgn0003944
C:Keywords: DNA binding; nucleus; transcription regulation

Query Match 25.0%; Score 2; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
||
DB 3 sy 4

RESULT 9
B57444
neuropeptide Grb-AST B2 - two-spotted cricket
C:Species: *Gryllus bimaculatus* (two-spotted cricket)
C:Date: 26-Jan-1996 #sequence_revision 26-Jan-1996 #text_change 26-Jan-1996
C:Accession: B57444
R:Lorenz, M.W.; Kollner, R.; Hoffmann, K.H.
J. Biol. Chem. 270, 21103-21108, 1995
A:Title: A family of neuropeptides that inhibit juvenile hormone biosynthesis in the cricket
A:Reference number: A57444; MUID:95403341
A:Accession: B57444
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-9 <LOR>

Query Match 25.0%; Score 2; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wr 2
||
DB 2 wr 3

RESULT 10
C57444
neuropeptide Grb-AST B3 - two-spotted cricket
C:Species: *Gryllus bimaculatus* (two-spotted cricket)
C:Date: 26-Jan-1996 #sequence_revision 26-Jan-1996 #text_change 26-Jan-1996
C:Accession: C57444
R:Lorenz, M.W.; Kollner, R.; Hoffmann, K.H.
J. Biol. Chem. 270, 21103-21108, 1995
A:Title: A family of neuropeptides that inhibit juvenile hormone biosynthesis in the cricket
A:Reference number: A57444; MUID:95403341
A:Accession: C57444
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-9 <LOR>

Query Match 25.0%; Score 2; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wr 2
||
DB 2 wr 3

RESULT 11
PT0238
Ig heavy chain CRD3 region (clone 2-94b) - human (fragment)
C:Species: *Homo sapiens* (man)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Aug-1996
C:Accession: PT0238
R:Tamada, M.; Wasserman, R.; Reichard, B.A.; Shane, S.; Caton, A.J.; Rovera, G.
J. Exp. Med. 173, 395-407, 1991
A:Title: Preferential utilization of specific immunoglobulin heavy chain diversity an
A:Reference number: PT0222; MUID:91108337
A:Accession: PT0238
A:Molecule type: DNA
A:Residues: 1-9 <YAM>
A:Experimental source: B lymphocyte
C:Keywords: heterotetramer; immunoglobulin

Query Match 25.0%; Score 2; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
||
DB 7 sy 8

RESULT 12
PT0324
Ig heavy chain CRD3 region (clone J2-106C) - human (fragment)
C:Species: *Homo sapiens* (man)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Aug-1996
C:Accession: PT0324
R:Yamada, M.; Wasserman, R.; Reichard, B.A.; Shane, S.; Caton, A.J.; Rovera, G.
J. Exp. Med. 173, 395-407, 1991
A:Title: Preferential utilization of specific immunoglobulin heavy chain diversity an
A:Reference number: PT0222; MUID:91108337
A:Accession: PT0324
A:Molecule type: DNA
A:Residues: 1-9 <YAM>
A:Experimental source: B lymphocyte
C:Keywords: heterotetramer; immunoglobulin

Query Match 25.0%; Score 2; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
||
DB 7 sy 8

RESULT 13
I46016
cytokeratin 4 - bovine (fragment)
C:Species: *Bos primigenius taurus* (cattle)
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 05-Nov-1999
C:Accession: I46016
R:Blessing, M.; Jorcano, J.L.; Franke, W.W.
EMBO J. 8, 117-126, 1989
A:Title: Enhancer elements directing cell-type-specific expression of cytokeratin gen
A:Reference number: I46016; MUID:89231609
A:Accession: I46016
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-9 <BLF>
A:Cross-references: EMBL:X14478; NID:9303; PIDN:CAA32640.1; PID:9577897

Query Match 25.0%; Score 2; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
||
Db 2 sy 3

RESULT 14

A28924
Fructose-bisphosphate aldolase (PC 4.1.2.13) B, hepatic - rabbit (fragment)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 28-Aug-1989 #sequence_revision 28-Aug-1989 #text_change 07-Feb-1997
C:Accession: A28924
R:Jacko, A.G.; Brox, L.W.; Gracy, R.W.; Horecker, B.L.
J. Biol. Chem. 245, 2140-2141, 1970
A:Title: The carboxyl-terminal structure of rabbit liver aldolase (aldolase B).
A:Reference number: A28924; MUID:70166720
C:Accession: A28924
A:Molecule type: protein
A:Residues: 1-9 <IAX>
C:Keywords: aldehyde-lyase; carbon-carbon lyase; gluconeogenesis; glycolysis; liver; pen

Query Match 25.0%; Score 2; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
||
Db 8 sy 9

RESULT 15

S36850
Ig heavy chain V region - mouse
C:Species: Mus musculus (house mouse)
C:Date: 06-Jan-1994 #sequence_revision 01-Dec-1995 #text_change 05-Nov-1999
C:Accession: S36850
R:Jacobs, J.; Kelsoe, G.
submitted to the EMBL Data Library, July 1992
A:Description: In situ studies on the primary immune response to (4-hydroxy-3-nitropheny
A:Reference number: S25024
A:Accession: S36850
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-9 <JAC>
A:Cross-references: EMBL:X67387; NID:950113; PIDN:CAA47799.1; PID:e51594; PID:g1333871
C:Keywords: heterotetramer; immunoglobulin

Query Match 25.0%; Score 2; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
||
Db 7 sy 8

RESULT 16

G41946
T-cell receptor gamma chain (2t.23) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 07-May-1999
C:Accession: G41946
R:Whetsell, M.; Mosley, R.L.; Whetsell, L.; Schaefer, F.V.; Miller, K.S.; Klein, J.R.
Mol. Cell. Biol. 11, 5902-5909, 1991
A:Title: Rearrangement and junctional-site sequence analyses of T-cell receptor gamma ge
A:Reference number: A41946; MUID:92049316
C:Accession: G41946
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-9 <WHB>
C:Keywords: T-cell receptor

Query Match 25.0%; Score 2; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
||
Db 1 sy 2

RESULT 17

RHPGG
gonadoliberin - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 18-Mar-1997
C:Accession: A01411
R:Baba, Y.; Matsuo, H.; Schally, A.V.
Biochem. Biophys. Res. Commun. 44, 459-463, 1971
A:Title: Structure of the porcine LH- and FSH-releasing hormone. II. Confirmation of
A:Reference number: A90172; MUID:72114303
A:Accession: A01411
A:Molecule type: protein
A:Residues: 1-10 <BAB>
R:Matsuo, H.; Arimura, A.; Nair, R.M.G.; Schally, A.V.
Biochem. Biophys. Res. Commun. 45, 822-827, 1971
A:Title: Synthesis of the porcine LH- and FSH-releasing hormone by the solid-phase me
A:Reference number: A90176; MUID:72065376
A:Contents: annotation; synthesis
A:Note: the synthetic and natural hormones have the same physicochemical and biologic
R:Baba, Y.; Arimura, A.; Schally, A.V.
Biochem. Biophys. Res. Commun. 45, 483-487, 1971
A:Title: On the tryptophan residue in porcine LH and FSH-releasing hormone.
A:Reference number: A90175; MUID:72117544
A:Contents: annotation
A:Note: Trp-3 appears to be essential for biological activity
C:Comment: This hypothalamic hormone stimulates the secretion of both luteinizing and
C:Superfamily: gonadoliberin
C:Keywords: amidated carboxyl end; hormone: hypothalamus: pyroglutamic acid
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F;10/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 25.0%; Score 2; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
||
Db 4 sy 5

RESULT 18

RHSHG
gonadoliberin - sheep
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 18-Mar-1997
C:Accession: A93780; A01411
R:Burgus, R.; Butcher, M.; Amoss, M.; Ling, N.; Monahan, M.; Rivier, J.; Fellows, R.;
Proc. Natl. Acad. Sci. U.S.A. 69, 278-282, 1972
A:Title: Primary structure of the ovine hypothalamic luteinizing hormone-releasing fa
A:Reference number: A93780; MUID:72094314
A:Accession: A93780
A:Molecule type: protein
A:Residues: 1-10 <BUR>
A:Note: the natural and synthetic hormones have the same biological activity
C:Comment: This hypothalamic hormone stimulates the secretion of both luteinizing and
C:Superfamily: gonadoliberin
C:Keywords: amidated carboxyl end; hormone: hypothalamus: pyroglutamic acid
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F;10/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 25.0%; Score 2; DB 1; Length 10;

```
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
   ||
Db 4 sy 5

RESULT 19
gonadoliberlin I - American alligator
N:Alternate names: gonadotropin-releasing hormone I
C:Species: Alligator mississippiensis (American alligator)
C>Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 18-Mar-1997
C:Accession: A60066
R:Lovejoy, D.A.; Fischer, W.H.; Parker, D.B.; McRory, J.E.; Park, M.; Lance, V.; Swanson
Regul. Pept. 33, 105-116, 1991
A:Title: Primary structure of two forms of gonadotropin-releasing hormone from brains of
A:Reference number: A60066; MUID:91352338
A:Accession: A60066
A:Molecule type: protein
A:Residues: 1-10 <LOV>
C:Superfamily: gonadoliberlin
C:Keywords: amidated carboxyl end; hormone; hypothalamus; pyroglutamic acid
F:1/Modified site: pyroglutamate carboxylic acid (Glu) #status experimental
F:10/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 25.0%; Score 2; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
   ||
Db 4 sy 5

RESULT 20
gonadoliberlin - chum salmon
C:Species: Oncorhynchus keta (chum salmon)
C>Date: 10-Aug-1990 #sequence_revision 10-Aug-1990 #text_change 18-Jun-1993
C:Accession: A21114
R:Sherwood, N.; Elden, L.; Brownstein, M.; Spiess, J.; Rivier, J.; Vale, W.
Proc. Natl. Acad. Sci. U.S.A. 80, 2794-2798, 1983
A:Title: Characterization of a teleost gonadotropin-releasing hormone.
A:Reference number: A21114; MUID:83195140
A:Accession: A21114
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-10 <SHE>

Query Match 25.0%; Score 2; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
   ||
Db 4 sy 5

RESULT 21
T-cell receptor beta chain V-D-J region - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C>Date: 09-Oct-1992 #sequence_revision 19-Oct-1995 #text_change 30-May-1997
C:Accession: PH08948; PH08947; PH0909; PH0899
R:Gold, D.P.; Offner, H.; Sun, D.; Wiley, S.; Vandenbark, A.A.; Wilson, D.B.
J. Exp. Med. 174, 1467-1476, 1991
A:Title: Analysis of T cell receptor beta chains in Lewis rats with experimental allergic
A:Reference number: PH0891; MUID:92078857
A:Accession: PH0948
```

```
A:Molecule type: mRNA
A:Residues: 1-10 <GO1>
A:Experimental source: myelin basic protein fragment-reactive T-cell, recovered from
A:Accession: PH0897
A:Molecule type: mRNA
A:Residues: 1-10 <GO2>
A:Experimental source: myelin basic protein-immunized T-cell, clones 3, 6-2, 14, hybr
A:Accession: PH0909
A:Molecule type: mRNA
A:Residues: 1-10 <GO3>
A:Accession: PH0899
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-10 <GO4>
A:Experimental source: clone 14
C:Keywords: T-cell receptor
```

```
Query Match 25.0%; Score 2; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 5 sy 6
   ||
Db 7 sy 8
```

```
RESULT 22
A40693
transgelin - sheep (fragment)
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C>Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 31-Oct-1997
C:Accession: A40693
R:Shapland, C.; Hsuan, J.J.; Totty, N.F.; Lawson, D.
J. Cell Biol. 121, 1065-1073, 1993
A:Title: Purification and properties of transgelin: a transformation and shape change
A:Reference number: A40693; MUID:93273790
A:Accession: A40693
A:Molecule type: protein
A:Residues: 1-11 <SHA>
A:Experimental source: aorta
C:Comment: This protein gels actin and is down regulated by transformation or loss of
C:Superfamily: smooth muscle protein SM22; calponin repeat homology; smooth muscle pr
C:Keywords: actin binding; cytoskeleton
```

```
Query Match 25.0%; Score 2; DB 2; Length 11;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 5 sy 6
   ||
Db 4 sy 5
```

```
RESULT 23
A34662
Achatina cardio-excitatory peptide-1 - giant African snail
C:Species: Achatina fulica (giant African snail)
C>Date: 22-Jun-1990 #sequence_revision 22-Jun-1990 #text_change 11-Jul-1997
C:Accession: A34662
R:Fujimoto, K.; Ohta, N.; Yoshida, M.; Kubota, I.; Muneoka, Y.; Kobayashi, M.
Biochem. Biophys. Res. Commun. 167, 777-783, 1990
A:Title: A novel cardio-excitatory peptide isolated from the atria of the African gila
A:Reference number: A34662; MUID:90211261
A:Accession: A34662
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-11 <FUU>
C:Keywords: amidated carboxyl end
F:11/Modified site: amidated carboxyl end (Phe) #status experimental
```

Query Match 25.0%; Score 2; DB 2; Length 11;
 Best Local Similarity 100.0%; Pred. No. 1.9e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wr 2
 ||
 DB 5 WR 6

RESULT 24

B41946
 T-cell receptor gamma chain (lt.57) - mouse (fragment)

C:Species: Mus musculus (house mouse)
 C:Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 07-May-1999

C:Accession: B41946

R:Whetsell, M.; Mosley, R.L.; Whetsell, L.; Schaefer, F.V.; Miller, K.S.; Klein, J.R.
 Mol. Cell. Biol. 11, 5902-5909, 1991

A:Title: Rearrangement and junctional-site sequence analyses of T-cell receptor gamma gene
 A:Reference number: M41946; MUID:92049316

A:Accession: B41946

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-11 <RES>

C:Keywords: T-cell receptor

Query Match 25.0%; Score 2; DB 2; Length 11;
 Best Local Similarity 100.0%; Pred. No. 1.9e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wr 2
 ||
 DB 5 WR 6

RESULT 25

I60434

68kDa neurofilament - rat (fragment)

C:Species: Rattus norvegicus (Norway rat)

C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999

C:Accession: I60434

R:Reeben, M.; Neuman, T.; Palgi, J.; Palm, K.; Paalme, V.; Saarna, M.
 J. Neurosci. Res. 40, 177-188, 1995

A:Title: Characterization of the rat light neurofilament (NF-L) gene promoter and identi
 A:Reference number: I60434; MUID:95264348

A:Accession: I60434

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-11 <RES>

A:Cross-references: EMBL:X53981; NID:g452676; PIDN:CAA37931.1; PID:g452677

C:Genetics:

A:Gene: NF68

Query Match 25.0%; Score 2; DB 2; Length 11;
 Best Local Similarity 100.0%; Pred. No. 1.9e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
 ||
 DB 5 sy 6

Search completed: January 14, 2002, 07:58:37
 Job time: 387 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 08:08:22 ; Search time 30.66 Seconds

(without alignments)
9.567 Million cell updates/sec

Title: 09-185908-1E

Perfect score: 8

Sequence: 1 wrxxsyxg 8

Scoring table: OLIGO

Searched: 100059 segs, 36664827 residues

Word size: 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 50 summaries

Database: SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25.0	7	1	HY7_PIG	P01153 sus scrofa
2	25.0	7	1	MMAL_ACHFU	P35919 aachalina fu
3	25.0	8	1	AL16_CARMA	P81819 carcinus ma
4	25.0	10	1	CON1_ALMT	P37041 alligator m
5	25.0	10	1	GON3_ONCKE	P20367 mus musculu
6	25.0	10	1	RLA2_MOUSE	P99027 mus musculu
7	25.0	11	1	CEP1_ACHFU	P22790 aachalina fu
8	25.0	12	1	CKA1_CONIM	P50983 conus imper
9	25.0	13	1	MLA_ANOCA	P41589 anolis caro
10	25.0	13	1	MLA_CAMDR	P01198 calotropis
11	25.0	14	1	CAL1_CALGI	P20728 calotropis
12	25.0	14	1	LPW_CITFR	P03056 citrobacter
13	25.0	14	1	LPW_ECOLI	P03053 escherichia
14	25.0	14	1	LPW_SALTY	P03054 salmonella
15	25.0	15	1	CKA2_CONAL	P36640 conus aulic
16	25.0	15	1	PH2_PERAM	P82695 periplaneta
17	25.0	15	1	UC08_MAIZE	P80614 zea mays (m
18	25.0	16	1	CAT9_FASHE	P80533 fasciola he
19	25.0	16	1	CKA1_CONAL	P56639 conus aulic
20	25.0	16	1	CKA1_CONAL	P56641 conus aulic
21	25.0	16	1	RIPK_TRIKI	P16093 trichosanthe
22	25.0	17	1	LPW_CORGL	P06556 coriaria he
23	25.0	17	1	PH3_PERAM	P82696 periplaneta
24	25.0	18	1	AGI_EUPCH	P33888 euphorbia c
25	25.0	18	1	AL12_CYPDO	P82153 cydia pomon
26	25.0	18	1	FMF1_ECOLI	P28660 escherichia
27	25.0	18	1	GALS_SALTY	P41030 salmonella
28	25.0	18	1	PHPT_PSESE	P23271 pseudocaria
29	25.0	19	1	CAT3_FASHE	P80532 fasciola he
30	25.0	19	1	COXR_THUOB	P80984 thunnus obe
31	25.0	19	1	DHAB_COMTE	P80704 comamonas t
32	25.0	19	1	DURC_STRGP	P35503 streptomyce
33	25.0	20	1	CIST_STNHY	P20903 streptomyce

34	2	25.0	20	1	CRP_MUSCA	P19094 mustelus ca
35	2	25.0	20	1	LPB3_HUMAN	P56643 homo sapien
36	2	25.0	20	1	PYR2_PYRAP	P37362 pyrriocoris
37	2	25.0	20	1	RIPX_CUCPE	P80750 cucurbita p
38	2	25.0	20	1	RLC1_HALMA	P12740 haloragula
39	2	25.0	20	1	VR90_BORPE	P81549 bordelella
40	2	25.0	21	1	FIBB_ANTAM	P14465 antilocapra
41	2	25.0	21	1	NRLA_ACISP	P33036 acinetobact
42	2	25.0	21	1	PSBF_SYNVU	P12239 synecococc
43	2	25.0	22	1	CYSP_TRIVA	P33404 trichomonas
44	2	25.0	22	1	LANW_SRRMU	P80666 streptococc
45	2	25.0	22	1	TX1_HERFU	P82850 heterometru
46	2	25.0	23	1	CYSP_TRIFO	P33403 tritrichomo
47	2	25.0	23	1	TX2_HERFU	P82851 heterometru
48	2	25.0	24	1	CS33_ARAHY	P80926 arachis hyp
49	2	25.0	24	1	FEDG_AMEYE	P80707 amycolatops
50	2	25.0	24	1	FIBG_CANFA	P12800 canis famli

ALIGNMENTS

RESULT 1
HY7_PIG
ID HY7_PIG STANDARD: PRT: 7 AA.
AC P01153;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 21-JUL-1986 (Rel. 01, Last annotation update)
DE HYPOTHALAMIC HEPTRAPEPTIDE.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_Taxid=96823;
RN [1]
RP SEQUENCE, AND SYNTHESIS.
RX MEDLINE=81213980; PubMed=6263778;
RA Chang R.C.C., Huang W.-Y., Arimura A., Redding T.W., Coy D.H.,
SA Safran M., Kong A., Hamilton J.W., Cohn D.V., Schally A.V.;
RT "Isolation, structure and synthesis of a heptapeptide with in vitro
ACPH-releasing activity from porcine hypothalamus.";
RL Horm. Metab. Res. 13:228-232(1981).
DR PIR: A01417; NYP67.
SQ SEQUENCE 7 AA: 957 MW: 632B45B1FB5059A0 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 7;
Best local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sy 6
||
Db 5 sy 6

RESULT 2
MMAL_ACHFU
ID MMAL_ACHFU STANDARD: PRT: 7 AA.
AC P35919;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE MMAMIDE-1.
OS Aachalina fulica (Giant African snail).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;
OC Aachalinacea; Aachalinidae; Aachalina.
OX NCBI_Taxid=6530;
RN [1]
RP SEQUENCE.
RX MEDLINE=93265912; PubMed=8495720;
RA Minakata H., Ikeda T., Muneoka Y., Kobayashi M., Nomoto K.;
RT "Mwamide-1, -2 and -3: novel neuromodulatory peptides isolated from

```

RT ganglia of the African giant snail, Achatina fulica.";
RL FEBS Lett. 323:104-108(1993).
CC -I- FUNCTION: EXHIBITS MODULATORY EFFECTS ON THE PERIPHERAL NERVOUS
CC SYSTEM. INHIBITS ACTIVITY ON A CENTRAL NEURON.
DR PIR. S33245; S33245.
KW Neuropeptide; Amidation.
FT MOD_RES 7 AA; 993 MW; 7362D5B69B041310 CRC64;
SQ SEQUENCE 7 AA; 993 MW; 7362D5B69B041310 CRC64;

Query Match
Best Local Similarity 100.0%; Score 2; DB 1; Length 7;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wr 2
   11
Db 1 wr 2

RESULT 3
AL16_CARMA STANDARD; PRT; 8 AA.
ID AL16_CARMA
AC P81819;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CARCINOSTATIN 16.
OS Carcinus maenas (Common shore crab) (Green crab).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC Eubrachyura; Portunoidae; Portunidae; Carcinus.
NCBI_TaxID=6759;
RN NCB1
RP SEQUENCE.
RX TISSUE=Cerebral ganglion, and thoracic ganglion;
RX MEDLINE=98121193; PubMed=9461295;
RX Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
RX Thorpe A.;
RA "Isolation and identification of multiple neuropeptides of the
RT allatostatin superfamily in the shore crab Carcinus maenas.";
RL Eur. J. Biochem. 250:727-734(1997).
CC -I- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.
CC -I- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW Neuropeptide; Amidation; Multigene family.
FT MOD_RES 8 AA; 813 MW; 7C286B45BA476878 CRC64;
SQ SEQUENCE 8 AA; 813 MW; 7C286B45BA476878 CRC64;

Query Match
Best Local Similarity 100.0%; Score 2; DB 1; Length 8;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
   11
Db 5 sy 6

RESULT 4
GON1_AL1MI STANDARD; PRT; 10 AA.
ID GON1_AL1MI
AC P37041; P20407;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GONADOLIBERIN I (GONADOTROPIN-RELEASING HORMONE I) (GNRH-I) (LH-RH I)
DE (LUTIBERIN I).
OS Alligator mississippiensis (American alligator).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Crocodylidae; Alligatorinae; Alligator.
NCBI_TaxID=8496;
RN NCB1
RP SEQUENCE.
RX TISSUE=Brain;

```

```

RX MEDLINE=91352338; PubMed=1882082;
RA Lovejoy D.A., Fischer W.H., Parker D.B., McRory J.E., Park M.,
RA Lance V., Swanson P., Rivier J.E., Sherwood N.M.;
RT "Primary structure of two forms of gonadotropin-releasing hormone
RT from brains of the American alligator (Alligator mississippiensis).";
RL Regul. Pept. 33:105-116(1991).
CC -I- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS.
CC -I- SIMILARITY: BELONGS TO THE GNRH FAMILY.
DR PIR. A60066; RHA01.
DR InterPro: IPR002012; GNRH.
DR Pfam: PF00446; GNRH; 1.
DR PROSITE: PS00473; GNRH; 1.
KW Hormone; Amidation; Hypothalamus.
FT MOD_RES 10 AA; 1172 MW; 284B23D7286B45A3 CRC64;
SQ SEQUENCE 10 AA; 1172 MW; 284B23D7286B45A3 CRC64;

Query Match
Best Local Similarity 100.0%; Score 2; DB 1; Length 10;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
   11
Db 4 sy 5

RESULT 5
GON3_ONCKE STANDARD; PRT; 10 AA.
ID GON3_ONCKE
AC P20367; P81751;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GONADOLIBERIN II (GONADOTROPIN-RELEASING HORMONE II) (GNRH-II) (LH-
DE RH II) (LUTIBERIN II).
GN GNRH3.
OS Oncorhynchus keta (Chum salmon), and
OS Clupea pallasii (Pacific herring).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Plectanctopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
NCBI_TaxID=8018, 30724;
RN NCB1
RP SEQUENCE.
RX SPECIES=O. keta;
RX MEDLINE=83195140; PubMed=6341999;
RX Sherwood N., Elden L., Brownstein M., Spiess J., Rivier J., Vale W.;
RT "Characterization of a teleost gonadotropin-releasing hormone.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:2794-2798(1983).
RN NCB1
RP SEQUENCE, AND FUNCTION.
RX SPECIES=C. pallasii; TISSUE=Brain, and Pituitary;
RX MEDLINE=20114351; PubMed=10650929;
RX Carlsfeld J., Powell J.F.F., Park M., Fischer W.H., Craig A.G.,
RX Chang J.P., Rivier J.E., Sherwood N.M.;
RT "Primary structure and function of three gonadotropin-releasing
RT hormones, including a novel form, from an ancient teleost, herring.";
RL Endocrinology 141:505-512(2000).
CC -I- FUNCTION: STIMULATES THE SECRETION OF BOTH LUTEALIZING AND
CC FOLLICLE-STIMULATING HORMONES.
CC -I- SIMILARITY: BELONGS TO THE GNRH FAMILY.
DR PIR. A21114; A21114.
DR InterPro: IPR002012; GNRH.
DR Pfam: PF00446; GNRH; 1.
DR PROSITE: PS00473; GNRH; 1.
KW Hormone; Amidation; Hypothalamus.
FT MOD_RES 10 AA; 1230 MW; 284B23D7286B45A3 CRC64;
SQ SEQUENCE 10 AA; 1230 MW; 284B23D7286B45A3 CRC64;

Query Match
Best Local Similarity 100.0%; Score 2; DB 1; Length 10;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```


Best Local Similarity 100.0%; Pred. No. 9.6e+02;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sy 6
11
DB 4 sy 5

RESULT 6
ID RLA2_MOUSE
AC P99027;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 60S ACIDIC RIBOSOMAL PROTEIN P2 (FRAGMENT).
GN RLP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
NCBI_TaxID=10090;
RN 11
RP SEQUENCE.
RC TISSUE-Liver;
RA Sanchez J.-C., Rouge V., Frutiger S., Hughes G., Yan J.X.,
Hoosland C., Appel R.D., Binz P.-A., Hochstrasser D.F.,
Cowthorne M.,
Submitted (Aug-1998) to the SWISS-PROT data bank.
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE ELONGATION STEP OF
PROTEIN SYNTHESIS.
CC -1- SUBUNIT: P1 AND P2 EXIST AS DIMERS AT THE LARGE RIBOSOMAL
SUBUNIT.
CC -1- PTM: PHOSPHORYLATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
DR SWISS-2DPAGE: P99027; MOUSE.
KW Ribosomal protein; Phosphorylation.
FT NON_TER 10 10
SQ SEQUENCE 10 AA: 1186 MW; 07121E3B45BD2DB CRC64;

Query Match 25.0%; Score 2; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 9.6e+02;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sy 6
11
DB 6 sy 7

RESULT 7
ID CEPI_ACHFV
AC P22790;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE CARDIO-EXCITATORY PEPTIDE-1 (ACEP-1).
OS Achatina fulica (Giant African snail).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;
Achatinacea; Achatinidae; Achatina.
NCBI_TaxID=6530;
RN 11
RP SEQUENCE.
RC STRAIN-FERUSSAC; TISSUE-Heart atrium;
MEDLINE=90211261; PubMed=2322251;
RA Fujimoto K., Ohta N., Yoshida M., Kubota I., Munoeke Y., Kobayashi M.,
"A novel cardio-excitatory peptide isolated from the atria of the
African giant snail, Achatina fulica.";
RT Biochem. Biophys. Res. Commun. 167:777-783(1990).
CC -1- FUNCTION: POTENTIATES THE BEAT OF THE VENTRICLE, AND HAS ALSO
EXCITATORY ACTIONS ON THE PENIS RETRACTOR MUSCLE, THE BUCCAL
MUSCLE AND THE IDENTIFIED NEURONS CONTROLLING THE BUCCAL MUSCLE
MOVEMENT OF ACHATINA.

CC -1- SIMILARITY: TO POSSIBLE PEPTIDE L5 FROM APLYSIA.
DR PIR: A34662; A34662.
KW Hormone; Amidation.
FT MOD_RES 11
SQ SEQUENCE 11 AA: 1305 MW; 82D6D5B9C7741365 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 11;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wr 2
11
DB 5 wr 6

RESULT 8
ID CXAL_CONIM
AC P50983;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ALPHA-CONOTOXIN IMI.
OS Conus imperialis (Imperial cone).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Caenogastropoda;
Neogastropoda; Conoidea; Conidae; Conus.
NCBI_TaxID=35631;
RN 11
RP SEQUENCE, AND SYNTHESIS.
RC TISSUE-Venom;
RX MEDLINE=94266889; PubMed=8206995;
RA McIntosh J.M., Yoshikami D., Mahe E., Nielsen D.B., Rivier J.E.,
Gray W.R., Olivera B.M.,
"A nicotinic acetylcholine receptor ligand of unique specificity,
RT alpha-conotoxin ImI.";
RT J. Biol. Chem. 269:16733-16739(1994).
RN 12
RP CHARACTERIZATION.
RX MEDLINE=95379776; PubMed=7651351;
RA Johnson D.S., Martinez J., Elgoyhen A.B., Heinemann S.F.,
McIntosh J.M.,
"Alpha-conotoxin ImI exhibits subtype-specific nicotinic
RT acetylcholine receptor blockade: preferential inhibition of homomeric
RT alpha 7 and alpha 9 receptors.";
RT Mol. Pharmacol. 48:194-199(1995).
RN 13
RP STRUCTURE BY NMR.
RX MEDLINE=99212205; PubMed=10194298;
RA Rogers J.P., Lugnbuhl P., Shen G.S., McCabe R.T., Stevens R.C.,
Wemmer D.E.,
"NMR solution structure of alpha-conotoxin ImI and comparison to
RT other conotoxins specific for neuronal nicotinic acetylcholine
RT receptors.";
RN Biochemistry 38:3874-3882(1999).
RN 14
RP STRUCTURE BY NMR.
RX MEDLINE=99158061; PubMed=10050774;
RA Maslennikov I.V., Shenkarev Z.O., Zhmak M.N., Ivanov V.T.,
Mehrfessel C., Tselin V.I., Arseniev A.S.,
"NMR spatial structure of alpha-conotoxin ImI reveals a common
RT scaffold in snail and snake toxins recognizing neuronal nicotinic
RT acetylcholine receptors.";
RN FEBS Lett. 444:275-280(1999).
RN 15
RP STRUCTURE BY NMR.
RX MEDLINE=99324017; PubMed=10395477;
RA Gehrmann J., Daly N.L., Alewood P.F., Craik D.J.,
"Solution structure of alpha-conotoxin ImI by 1H nuclear magnetic
RT resonance.";
RN J. Med. Chem. 42:2364-2372(1999).
CC -1- FUNCTION: ALPHA-CONOTOXIN ACT ON POSTSYNAPTIC MEMBRANES, THEY
BIND TO THE NICOTINIC ACETYLCHOLINE RECEPTORS (NACHR) AND THUS

```

CC INHIBIT THEM. IT IS HIGHLY ACTIVE AGAINST THE NEUROMUSCULAR
CC RECEPTOR IN FROG BUT NOT IN MICE. IN CONTRAST, IT INDUCES SEIZURES
CC WHEN INJECTED CENTRALLY IN MICE AND RATS. IT TARGET NEURONAL
CC NMDARS IN MAMMALS. BLOCKS HOMOMERIC ALPHA-7 NICOTINIC RECEPTORS
CC WITH THE HIGHEST APPARENT AFFINITY AND HOMOMERIC ALPHA-9 RECEPTORS
CC WITH 8-FOLD LOWER AFFINITY. HAS NO EFFECT ON RECEPTORS COMPOSED OF
CC ALPHA-2/BETA-2, ALPHA-3/BETA-2, ALPHA-4/BETA-2, ALPHA-2/BETA-4,
CC ALPHA-3/BETA-4, OR ALPHA-4/BETA-4 SUBUNIT COMBINATIONS.
CC PDB: 11M1; 15-JUN-99.
DR PDB: 11M1; 23-APR-99.
DR PDB: 1CNL; 27-MAY-99.
KW Postsynaptic neurotoxin; Acetylcholine receptor inhibitor; Amidation;
KW Venom; 3D-structure.
FT DISULFID 2 2
FT DISULFID 3 12
FT MOD_RES 12 12
FT SEQUENCE 12 AA; 1357 MW; 9C29CEB545A4176A CRC64;

Query Match 25.0%; Score 2; DB 1; Length 12;
Best Local Similarity 100.0%; Pred No. 1.1e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wr 2
DB 10 wr 11

RESULT 9
MLA_ANOCA STANDARD; PRT; 13 AA.
AC P41589;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MELANOTROPIN ALPHA (ALPHA-MSH).
OS Anolis carolinensis (Green anole) (American chameleon).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Lepidosauria; Squamata; Iguania; Iguanidae; Polychrotinae; Anolis.
OX NCBI_TaxID=28377;
RN [1]
RP SEQUENCE.
RP TISSUE=pituitary;
RX MEDLINE=92270473; PubMed=1667689;
RA Dorez R.M., Lancha A., Rand-Weaver M., Jankelow L., Adamczyk D.L.;
RT "Detection of a novel sequence change in the major form of alpha-MSH
RT isolated from the intermediate pituitary of the reptile, Anolis
RT carolinensis."
RL Peptides 12:1261-1266(1991).
RL -1- SIMILARITY: BELONGS TO THE POMC FAMILY.
KW Hormone; Amidation.
FT MOD_RES 13
FT SEQUENCE 13 AA; 1608 MW; FP990A7358BB09C1 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 13;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sy 6
DB 1 sy 2

RESULT 10
MLA_CAMDR STANDARD; PRT; 13 AA.
AC P01198;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MELANOTROPIN ALPHA (ALPHA-MSH).
OS Camelus, dromedarius (Arabian camel), and

```

```

OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
OX NCBI_TaxID=9838; 9796;
RN [1]
RP SEQUENCE.
RC SPECIES=C.dromedarius;
RX MEDLINE=7516434; PubMed=1125179;
RA Li C.H., Danho W.O., Chung D., Rao A.J.;
RT "Isolation, characterization, and amino acid sequence of
RT melanotropins from camel pituitary glands."
RL Biochemistry 14:947-952(1975).
RN [2]
RP SEQUENCE.
RC SPECIES=Horse; TISSUE=pituitary;
RA Dixon J.S., Li C.H.;
RT "The isolation and structure of alpha-melanocyte-stimulating hormone
RT from horse pituitaries."
RL J. Am. Chem. Soc. 82:4568-4572(1960).
CC -1- SIMILARITY: BELONGS TO THE POMC FAMILY.
DR PIR: A01464; MTCMAD.
DR PIR: A91785; MTCMAD.
KW Hormone; Acetylation; Amidation.
FT MOD_RES 1
FT MOD_RES 1
FT MOD_RES 13
FT MOD_RES 13
FT SEQUENCE 13 AA; 1624 MW; FF991CA958BB09C1 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 13;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sy 6
DB 1 sy 2

RESULT 11
CALI_CALGI STANDARD; PRT; 14 AA.
AC P20728;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE CALOTROPIN DI (EC 3.4.22.-) (FRAGMENT).
OS Calotropis gigantea (Madar) (Bowstring hemp).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; easterids I; Gentianales; Apocynaceae; Asclepiadoideae;
OC Asclepiadeae; Calotropis.
OX NCBI_TaxID=4066;
RN [1]
RP SEQUENCE.
RA Bhattacharya D., Sengupta A., Sinha N.K.;
RT "Chemical modification and amino terminal sequence of calotropin DI
RT from Calotropis gigantea."
RL Phytochemistry 26:633-636(1987).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE
CC PAPAIN FAMILY OR THIOL PROTEASES.
DR PIR: PT0026; PT0026.
DR MEROPS: C01.011; -.
DR InterPro: IPR000169; Thiolprot_act_site.
DR PROSITE: PS00139; THIOL_PROTEASE_CYS; PARTIAL.
DR PROSITE: PS00639; THIOL_PROTEASE_HIS; PARTIAL.
DR PROSITE: PS00640; THIOL_PROTEASE_ASN; PARTIAL.
KW Hydrolyase; Thiol protease.
FT MOD_RES 1
FT MOD_RES 1
FT NON_TER 14
FT SEQUENCE 14 AA; 1715 MW; D933F0276CDA4662 CRC64;

Query Match 25.0%; Score 2; DB 1; Length 14;

```

Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wr 2
11
DB 8 WR 9

RESULT 12

LPW_CITFR STANDARD; PRT; 14 AA.
AC P03056;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-MAY-1991 (Rel. 18, Last annotation update)
GN TRP OPERON LEADER PEPTIDE.
DE TRPL.

OS Citrobacter freundii.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Citrobacter.

NCBI_TaxID=546;
[1]

SEQUENCE FROM N.A.
RX MEDLINE=83007061; PubMed=6749821;
RA Blumenberg M., Yanofsky C.;

"Evolutionary divergence of the Citrobacter freundii tryptophan
operon regulatory region: comparison with other enteric bacteria."
RL J. Bacteriol. 152:57-62(1982).

-I- FUNCTION: THIS PROTEIN IS INVOLVED IN CONTROL OF THE BIOSYNTHESIS
OF TRYPTOPHAN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: J01557; -; NOT_ANNOTATED_CDS.
DR PIR: A03592; LFEHWC.

TRYPTOPHAN biosynthesis; Leader peptide.
KW Tryptophan biosynthesis; Leader peptide.
SO SEQUENCE 14 AA: 1720 MW: 58792A473B804BE7 CRC64:

Query Match 25.0%; Score 2; DB 1; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wr 2
11
DB 11 WR 12

RESULT 13

LPW_ECOLI STANDARD; PRT; 14 AA.
AC P03053;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
GN TRP OPERON LEADER PEPTIDE.
DE TRPL OR TRPEE OR B1265 OR Z2545 OR ECS1837.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Escherichia.

NCBI_TaxID=562, 83334;
[1]

SEQUENCE FROM N.A.
RX MEDLINE=82150258; PubMed=7038627;
RA Yanofsky C., Platt T., Crawford I.P., Nichols B.P., Christie G.E.,
RA Horowitz H., van Cleemput M., Wu A.M.;

RT "The complete nucleotide sequence of the tryptophan operon of
RT Escherichia coli.";
RL Nucleic Acids Res. 9:6647-6668(1981).

RP SEQUENCE FROM N.A.
RX MEDLINE=76240562; PubMed=781271;
RA Squires C., Lee F., Bertrand K., Squires C.L., Bronson M.J.,
RA Yanofsky C.;
RT "Nucleotide sequence of the 5' end of tryptophan messenger RNA of
RT Escherichia coli.";
RL J. Mol. Biol. 103:351-381(1976).

RP SEQUENCE FROM N.A.
RX MEDLINE=80101455; PubMed=118451;
RA Oender D.L., Zurawski G., Yanofsky C.;
RT "Attenuation in the Escherichia coli tryptophan operon: role of RNA
RT secondary structure involving the tryptophan codon region.";
RL Proc. Natl. Acad. Sci. U.S.A. 76:5524-5528(1979).

RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).

RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamosis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).

RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).

-I- FUNCTION: THIS PROTEIN IS INVOLVED IN CONTROL OF THE BIOSYNTHESIS
OF TRYPTOPHAN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: J01714; AA57296.1; -;
DR EMBL: A04494; CA000361.1; -;
DR EMBL: AE000224; AAC74347.1; -;
DR EMBL: AE0005380; AAG56550.1; -;
DR EMBL: AP002556; BAB35260.1; -;
DR PIR: A03589; LFEHWC.
DR EcoGene: EG11274; trpl.
KW Tryptophan biosynthesis; Leader peptide; Complete proteome.
SO SEQUENCE 14 AA: 1723 MW: 5879306E3E804A37 CRC64;

```

Query Match      25.0%; Score 2; DB 1; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wr 2
   ||
Db 11 wr 12

RESULT 14
LPM_SALTY STANDARD; PRT; 14 AA.
ID LPM_SALTY
AC P03054;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE TRP OPERON LEADER PEPTIDE.
OS Salmoneella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmoneella.
OX NCBI_TaxID=602;

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=78196931; PubMed=351195;
RA Lee F., Bertrand K., Bennett G.N., Yanofsky C.;
RT "Comparison of the nucleotide sequences of the initial transcribed
RT regions of the tryptophan operons of Escherichia coli and Salmoneella
RT typhimurium."
RL J. Mol. Biol. 121:193-217(1978).
CC -1- FUNCTION: THIS PROTEIN IS INVOLVED IN CONTROL OF THE BIOSYNTHESIS
CC OF TRYPTOPHAN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; M24960; -; NOT_ANNOTATED_CDS.
DR PIR; A03590; LFEWBT.
DR StyGene; SG10400; trlp.
KM Tryptophan biosynthesis: Leader peptide.
SQ SEQUENCE 14 AA; 1635 MW; 49f22A47362248E7 CRC64;

Query Match      25.0%; Score 2; DB 1; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wr 2
   ||
Db 11 wr 12

RESULT 15
CXA2_CONAL STANDARD; PRT; 15 AA.
ID CXA2_CONAL
AC P56640;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE ALPHA-CONOTOXIN AUIB.
OS Conus aulicus (Court cone).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Caenogastropoda;
OC Neogastropoda; Conoidea; Conidae; Conus.
OX NCBI_TaxID=89437;
RN [1]
RP SEQUENCE, AND SYNTHESIS.
RC TISSUE=Venom;
RX MEDLINE=99003392; PubMed=9786965;

```

```

RA Luo S., Kulak J.M., Cartier G.E., Jacobsen R.B., Yoshikami D.,
RA Olivera B.M., McIntosh J.M.;
RT "Alpha-conotoxin AUIB selectively blocks alpha3 beta4 nicotinic
RT acetylcholine receptors and nicotine-evoked norepinephrine release.";
RL J. Neurosci. 18:8571-8579(1998).
CC -1- FUNCTION: ALPHA-CONOTOXINS ACT ON POSTSYNAPTIC MEMBRANES, THEY
CC BIND TO THE NICOTINIC ACETYLCHOLINE RECEPTORS (NACHR) AND THUS
CC INHIBIT THEM. THIS PEPTIDE BLOCKS MAMMALIAN NICOTINIC
CC ACETYLCHOLINE RECEPTORS COMPOSED OF ALPHA-3/BETA-4 SUBUNITS.
CC -1- MASS SPECTROMETRY: MM=1572.5; METHOD=ELECTROSPRAY.
KM Postsynaptic neurotoxin; Acetylcholine receptor inhibitor; Amidation;
KW Venom.
FT DISULFID 2 8
FT DISULFID 3 15
FT MOD_RES 15 15
SQ SEQUENCE 15 AA; 1578 MW; 84FEF95FDC700155 CRC64;

Query Match      25.0%; Score 2; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sy 6
   ||
Db 4 sy 5

RESULT 16
PH2_PERAM STANDARD; PRT; 15 AA.
ID PH2_PERAM
AC P82695;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PEPTIDE HORMONE 2 (PEA-VEAACID 2).
OS Periplaneta americana (American cockroach).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC Blattodea; Blattidae; Periplaneta.
OX NCBI_TaxID=6978;
RN [1]
RP SEQUENCE.
RC TISSUE=Abdominal perisymphatic organs;
RA Predel R.;
RL Submitted (JUL-2000) to the SWISS-PROT data bank.
CC -1- FUNCTION: UNKNOWN.
KW Neuropeptide.
SQ SEQUENCE 15 AA; 1603 MW; F353DC8B1F92B8BD CRC64;

Query Match      25.0%; Score 2; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 sy 6
   ||
Db 11 sy 12

RESULT 17
UC08_MAIZE STANDARD; PRT; 15 AA.
ID UC08_MAIZE
AC P80614;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE UNKNOWN PROTEIN FROM 2D-PAGE OF ETIOLOGED COLEOPTILE (SPOT 159)
DE (FRAGMENT).
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;

```

```

RN [1] *
RP SEQUENCE.
RC TISSUE-Coleoptile:
RA Teouzet P., Riccardi F., Morin C., Damerval C., Huet J.-C.,
RT Pernollet J.-C., Zivy M., de Vienne D.;
RT "The maize two dimensional gel protein database: towards an integrated
RT genome analysis program.";
RL Theor. Appl. Genet. 93:997-1005(1996).
CC -1- MISCELLANEOUS: ON THE 2D-GEL, THE DETERMINED PI OF THIS UNKNOWN
CC PROTEIN IS: 6.4, ITS MW IS: 38.8 KDA.
CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY.
DR Maize-2DPAGE: P80614; COLEOPTILE.
DR MaizeDB: 123934; -.
FT NON_TER 1 1
FT SEQUENCE 15 AA; 1785 MW; 1978B1D6A4DDF8D CRC64;

```

```

Query Match 25.0%; Score 2; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 wr 2
   11
DB 3 WR 4

```

```

RESULT 18
CAT9_FASHE STANDARD: PRT; 16 AA.
ID CAT9_FASHE P80533;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PUTATIVE CATHEPSIN-LIKE ENZYME (EC 3.4.22.-) (NEWLY EXCYSTED JUVENILE
DE PROTEIN 9) (FRAGMENT).
OS Fasciola hepatica (Liver fluke).
OS Eukaryota; Metazoa; Platyhelminthes; Turbellarian platyhelminths;
OC Rhadilophora; Eulielithophora; Reversoformata; Mediofusata;
OC Neodermata; Trematoda; Digenea; Echinostomida; Echinostomata;
OC Fasciolidae; Fasciolidae; Fasciola.
OX NCBI_TaxID=6192;
RN [1]
RP SEQUENCE.
RX MEDLINE=95366993; PubMed=7639732;
RA Tkalcic J., Ashman K., Moeusen E.;
RT "Fasciola hepatica: rapid identification of newly excysted juvenile
RT proteins.";
RL Biochem. Biophys. Res. Commun. 213:169-174(1995).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED AT THE NEWLY EXCYSTED JUVENILE
CC STAGE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1; ALSO KNOWN AS THE
CC PAPAIN FAMILY OF THIOL PROTEASES.
DR InterPro: IPR000169; ThiolProt_act_site.
DR PROSITE: PS00139; THIOL_PROTEASE_CYS; PARTIAL.
DR PROSITE: PS00639; THIOL_PROTEASE_HIS; PARTIAL.
DR PROSITE: PS00640; THIOL_PROTEASE_ASN; PARTIAL.
DR Hydrolyase; Thiol protease.
FT NON_TER 16 16
FT SEQUENCE 16 AA; 1966 MW; 087B18FDB1FA541E CRC64;

```

```

Query Match 25.0%; Score 2; DB 1; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 wr 2
   11
DB 8 WR 9

```

```

RESULT 19
ID CXAL_CONAL STANDARD: PRT; 16 AA.
AC P56639;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE ALPHA-CONOTOXIN AUIA.
OS Conus aulicus (Court cone).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Caenogastropoda;
OC Neogastropoda; Conoidea; Conidae; Conus.
OX NCBI_TaxID=89437;
RN [1]
RP SEQUENCE, AND SYNTHESIS.
RC TISSUE-Venom;
RX MEDLINE=99003392; PubMed=9786965;
RX Luo S., Kulak J.M., Cartier G.E., Jacobsen R.B., Yoshikami D.,
RA Oliveira B.M., McIntosh J.M.;
RT "Alpha-conotoxin AUIB selectively blocks alpha3 beta4 nicotinic
RT acetylcholine receptors and nicotine-evoked norepinephrine release.";
RL J. Neurosci. 18:8571-8579(1998).
CC -1- FUNCTION: ALPHA-CONOTOXINS ACT ON POSTSYNAPTIC MEMBRANES, THEY
CC BIND TO THE NICOTINIC ACETYLCHOLINE RECEPTORS (NACHR) AND THUS
CC INHIBIT THEM. THIS PEPTIDE BLOCKS MAMMALIAN NICOTINIC
CC ACETYLCHOLINE RECEPTORS COMPOSED OF ALPHA-3/BETA-4 SUBUNITS.
CC -1- MASS SPECTROMETRY: MW=1725.6; METHOD=ELECTROSPRAY.
KW Postsynaptic neurotoxin; Acetylcholine receptor inhibitor; Amidation;
KW Venom.
FT DISULFID 2 8
FT DISULFID 3 16
FT MOD_RES 16 16
FT SEQUENCE 16 AA; 1731 MW; 1E310FEB8FDC7001 CRC64;

```

```

Query Match 25.0%; Score 2; DB 1; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 sy 6
   11
DB 4 SY 5

```

```

RESULT 20
ID CXA3_CONAL STANDARD: PRT; 16 AA.
AC P56641;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE ALPHA-CONOTOXIN AUIA.
OS Conus aulicus (Court cone).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Caenogastropoda;
OC Neogastropoda; Conoidea; Conidae; Conus.
OX NCBI_TaxID=89437;
RN [1]
RP SEQUENCE, AND SYNTHESIS.
RC TISSUE-Venom;
RX MEDLINE=99003392; PubMed=9786965;
RX Luo S., Kulak J.M., Cartier G.E., Jacobsen R.B., Yoshikami D.,
RA Oliveira B.M., McIntosh J.M.;
RT "Alpha-conotoxin AUIB selectively blocks alpha3 beta4 nicotinic
RT acetylcholine receptors and nicotine-evoked norepinephrine release.";
RL J. Neurosci. 18:8571-8579(1998).
CC -1- FUNCTION: ALPHA-CONOTOXINS ACT ON POSTSYNAPTIC MEMBRANES, THEY
CC BIND TO THE NICOTINIC ACETYLCHOLINE RECEPTORS (NACHR) AND THUS
CC INHIBIT THEM. THIS PEPTIDE BLOCKS MAMMALIAN NICOTINIC
CC ACETYLCHOLINE RECEPTORS COMPOSED OF ALPHA-3/BETA-4 SUBUNITS.
CC -1- MASS SPECTROMETRY: MW=1667.6; METHOD=ELECTROSPRAY.
KW Postsynaptic neurotoxin; Acetylcholine receptor inhibitor; Amidation;
KW Venom.
FT DISULFID 2 8
FT DISULFID 3 16

```

```

FT MOD_RES 16 16 AMIDATION.
SQ SEQUENCE 16 AA: 1673 MW: 1E310D3B8FDC7001 CRC64;

Query Match          25.0%; Score 2; DB 1; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SY 6
   ||
Db 4 SY 5

RESULT 21
RIPK_TRIKI
ID RIPK_TRIKI STANDARD; PRT; 16 AA.
AC P16093;
DC 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE RIBOSOME-INACTIVATING PROTEIN TRICHOKIRIN (RNA N-GLYCOSIDASE)
   (EC 3.2.2.22) (FRAGMENT).
OS Trichosanthes kirilowii (Mongolian snake-gourd).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Cucurbitales; Cucurbitaceae; Trichosanthes.
OX NCBI_TaxID=3677;
RN [1]
RP SEQUENCE.
RC TISSUE=Seed;
RA MEDLINE=89005108; PubMed=3262509;
RA Casellas P., Dussossoy D., Falasca A.L., Barbieri L.,
RA Guillemot J.C., Ferrara P., Bolgonesi A., Genini P., Stirpe F.;
RT "Trichokirin, a ribosome-inactivating protein from the seeds of
RT Trichokirin, a ribosome-inactivating protein from the seeds of
RT characterizations and use for preparation of immunotoxins.";
RL Eur. J. Biochem. 176:581-588(1988).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF THE N-GLYCOSIDIC BOND AT ONE
CC SPECIFIC ADENOSINE ON THE 28S RNA.
CC -1- PTM: GLYCOSYLATED.
CC -1- SIMILARITY: TO OTHER BACTERIAL AND PLANTS RIBOSOME-INACTIVATING
CC PROTEINS. BELONGS TO TYPE 1 RIP.
CC PIR: S01669; S01669.
DR InterPro: IPR001574; RIP.
DR PROSITE: PS00275; SHIGA_RICIN: PARTIAL.
DR Protein synthesis inhibitor; Hydrolase; Toxin; Glycoprotein.
FT NON_TER 16 16
SQ SEQUENCE 16 AA: 1605 MW: 5E268A7F345935A2 CRC64;

Query Match          25.0%; Score 2; DB 1; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SY 6
   ||
Db 13 SY 14

RESULT 22
LPM_CORGL
ID LPM_CORGL STANDARD; PRT; 17 AA.
AC P06556;
DC 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRP OPERON LEADER PEPTIDE.
GN TRPL.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Firmicutes; Actinobacteriae; Actinobacteriaceae;
OC Actinomycetales; Corynebacteriineae; Corynebacteriaceae;
OC Corynebacterium.
OX NCBI_TaxID=1718;

```

```

RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=87117512; PubMed=3808947;
RA Matsui K., Sano K., Ohtsubo E.;
RT "Complete nucleotide and deduced amino acid sequences of the
RT Brevibacterium lactofermentum tryptophan operon.";
RL Nucleic Acids Res. 14:10113-10114(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88032866; PubMed=3667535;
RA Matsui K., Miwa K., Sano K.;
RT "Two single-base-pair substitutions causing desensitization to
RT tryptophan feedback inhibition of anthranilate synthase and enhanced
RT expression of tryptophan genes of Brevibacterium lactofermentum.";
RL J. Bacteriol. 169:5330-5332(1987).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=87277409; PubMed=3609747;
RA Sano K., Matsui K.;
RT "Structure and function of the trp operon control regions of
RT Brevibacterium lactofermentum, a glutamic-acid-producing bacterium.";
RL Gene 53:191-200(1987).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13059 / AS019;
RX MEDLINE=91088299; PubMed=2263476;
RA Heery D.M., Dunican L.K.;
RT "Nucleotide sequence of the Corynebacterium glutamicum trpE gene.";
RL Nucleic Acids Res. 18:7138-7138(1990).
CC -1- FUNCTION: THIS PROTEIN IS INVOLVED IN CONTROL OF THE BIOSYNTHESIS
CC OF TRYPTOPHAN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X04960; CAA28622.1; -
DR EMBL: M17892; AAB59110.1; -
DR EMBL: M16663; -; NOT_ANNOTATED_CDS.
DR EMBL: X55994; CAA39466.1; -.
DR PIR: A29458; A29458.
DR PIR: A29834; A29834.
DR PIR: A24723; A24723.
DR PIR: S13087; S13087.
KW tryptophan biosynthesis; Leader peptide.
SQ SEQUENCE 17 AA: 2112 MW: 74C7E7924DAE56B CRC64;

Query Match          25.0%; Score 2; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WR 2
   ||
Db 14 WR 15

RESULT 23
PH3_PERAM
ID PH3_PERAM STANDARD; PRT; 17 AA.
AC P82696;
DC 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PEPTIDE HORMONE 3 (PEA-VEAACID 1).
OS Periplaneta americana (American cockroach).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC Blattodea; Blattidae; Periplaneta.

```

```
OX NCB1_TaxID=6978;
RN [1]
RP SEQUENCE, AND MASS SPECTROMETRY.
RX TISSUE=Abdominal perisymphathetic organs;
RX MEDLINE=20140865; PubMed=10676456;
RA Pregel R., Eckert M., Holman G.M.;
RT "The unique neuropeptide pattern in abdominal perisymphathetic organs
of insects.";
RN Ann. N.Y. Acad. Sci. 897:282-290(1999).
CC -1- FUNCTION: UNKNOWN.
CC CC MASS SPECTROMETRY: MW=1651.35; METHOD=MALDI.
KW Neuropeptide.
SQ SEQUENCE 17 AA: 1807 MW: 2374AC8B1F86E8EB CRC64:

Query Match 25.0%; Score 2; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
Db 13 sy 14

RESULT 24
AGL_EUPCH STANDARD: PRT; 18 AA.
ID AGL_EUPCH
AC P33888;
DT 01-FEB-1994 (Rcl. 28, Created)
DT 01-FEB-1994 (Rcl. 28, Last sequence update)
DT 01-FEB-1994 (Rcl. 28, Last annotation update)
DE LECTIN (FRAGMENT).
OS Euphorbia characias (Spurge).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Malpighiales; Euphorbiaceae; Euphorbia.
OX NCB1_TaxID 3991;
RN [1]
RN SEQUENCE.
RP TISSUE=Latex;
RX MEDLINE=93357266; PubMed=8353129;
RA Silipe F., Licastro F., Morini M.C., Parente A., Savino G.,
RA Abbondanza A., Bolognesi A., Falasca A.I., Rossi C.A.;
RT "Purification and partial characterization of a mitogenic lectin from
the latex of Euphorbia marginata.";
RL Blochm. Biophys. Acta 1158:33-39(1993).
CC -1- FUNCTION: LECTIN THAT BINDS GALACTOSE. GALACTOSE-CONTAINING SUGARS
AND GENTIOBIOSE. IT IS STRONGLY MITOGENIC FOR HUMAN T LYMPHOCYTES.
CC -1- SUBUNIT: HOMODIMER.
CC -1- PTM: N-GLYCOSYLATED.
CC -1- SIMILARITY: TO E. MARGINATA LECTIN.
CC PIR: S36121; S36121.
KW Lectin.
RN NON_TER 18
SQ SEQUENCE 18 AA: 1923 MW: C6F6A1A7B2AB124F CRC64:

Query Match 25.0%; Score 2; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
Db 3 sy 4

RESULT 25
AL2_CYPDPO STANDARD: PRT; 18 AA.
ID AL2_CYPDPO
AC P82153;
DT 30-MAY-2000 (Rcl. 39, Created)
DT 30-MAY-2000 (Rcl. 39, Last sequence update)
DT 30-MAY-2000 (Rcl. 39, Last annotation update)
```

```
DE CYDIASTATIN 2.
OS Cydia pomonella (Codling moth).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Tortricidae; Tortricidae; Olethreutinae; Cydia.
OX NCB1_TaxID=82600;
RN [1]
RN SEQUENCE.
RP TISSUE=Larva;
RX MEDLINE=98054539; PubMed=9392829;
RA Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Winstanley D.,
RA Davey M., East P.D., Thoiry A.;
RT "Lepidopteran peptides of the allatostatin superfamily.";
RL Peptides 18:1301-1309(1997).
CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
KW Neuropeptide; Amidation.
FT MOD_RES 18
SQ SEQUENCE 18 AA: 2109 MW: 8E6679C0CDF175C CRC64:
```

```
Query Match 25.0%; Score 2; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
Db 3 sy 4
```

Search completed: January 14, 2002, 08:08:23
Job time: 708 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 08:07:34 ; Search time 81.98 Seconds
(without alignments)
14.274 Million cell updates/sec

Title: 09-185908-1E
Perfect score: 8
Sequence: 1 wxxsxyxg 8

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 50 summaries

Database :

- 1: SPREMBL_17:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37.5	38	3	09UR79	09ur79 trichoderma
2	37.5	77	5	09GVM9	09gvm9 leishmania
3	37.5	91	6	09MZE1	09mze1 bos laurus
4	37.5	131	13	091947	091947 oryzias lat
5	37.5	136	5	09GMB2	09gmb2 leishmania
6	37.5	141	6	027953	027953 balenopter
7	37.5	226	4	09N274	09n274 homo sapien
8	37.5	485	5	09GIV3	09giv3 trypanosoma
9	37.5	707	8	09B852	09b852 schoenocceph
10	25.0	8	4	015898	015898 homo sapien
11	25.0	8	4	034909	034909 locusta mtg
12	25.0	8	10	09S824	09s824 spinacia ol
13	25.0	8	11	062721	062721 rattus norv
14	25.0	9	4	09BFE9	09bfe9 homo sapien
15	25.0	9	6	028093	028093 bos laurus
16	25.0	9	11	09QWT0	09qwt0 mus musculu
17	25.0	9	13	091A14	091a14 gallus gall
18	25.0	11	4	09UCA6	09uca6 homo sapien
19	25.0	11	7	077876	077876 oreochromis

20	25.0	11	7	077883	077883 oreochromis
21	25.0	11	7	077886	077886 oreochromis
22	25.0	11	7	077912	077912 oreochromis
23	25.0	11	8	032704	032704 nicotiana t
24	25.0	11	8	09G607	09g607 aphaniotis
25	25.0	11	8	09G359	09g359 japalura fl
26	25.0	11	11	09JLE6	09jle6 rattus norv
27	25.0	12	4	09UM28	09um28 homo sapien
28	25.0	12	4	09UC05	09uc05 homo sapien
29	25.0	12	7	077919	077919 pseudotroph
30	25.0	12	12	081139	081139 barley str1
31	25.0	13	4	09UEE3	09uee3 homo sapien
32	25.0	13	4	09NR93	09nr93 homo sapien
33	25.0	13	8	09XLI2	09xli2 bemisia tab
34	25.0	13	12	064813	064813 autographa
35	25.0	14	2	052220	052220 salmonella
36	25.0	14	2	054081	054081 saccharopol
37	25.0	14	2	09R506	09r506 burkholderi
38	25.0	14	2	09R696	09r696 hordeum vul
39	25.0	14	8	09MT61	09mt61 allium cepa
40	25.0	14	8	09MRV4	09mrv4 allium porr
41	25.0	14	8	09MRV1	09mrv1 allium satl
42	25.0	14	8	09MRT8	09mrt8 alooe vera.
43	25.0	14	8	09MRE6	09mre6 hordeum mur
44	25.0	14	10	P82433	P82433 nicotiana t
45	25.0	14	10	09FYT0	09fyt0 allium cepa
46	25.0	14	11	070599	070599 rattus norv
47	25.0	14	11	09Z0G5	09z0g5 mus musculu
48	25.0	14	12	067112	067112 influenza a
49	25.0	14	12	086616	086616 turkey heap
50	25.0	15	2	Q47893	Q47893 fremyella d

ALIGNMENTS

RESULT 1	
09UR79	PRELIMINARY; PRT; 38 AA.
ID 09UR79;	
AC 09UR79;	
DT 01-MAY-2000 (TREMBLrel. 13, Created)	
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)	
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)	
DE CELLOBIOHYDROLASE II CORE PROTEIN, CBH II Cp-3.2.1.91.	
OS Trichoderma reesei (Hypocrea jecorina).	
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;	
OC Hypocreales; Hypocreaceae; Hypocrea.	
OX NCBI_TaxID=51453;	
RN [1]	
RP SEQUENCE.	
RX MEDLINE=94250349; PubMed=8192865;	
RA Woodward J., Brown J.P., Evans B.R., Affholter K.A.;	
RT "Papain digestion of crude Trichoderma reesei cellulase: purification	
RL and properties of cellobiohydrolase I and II core proteins.";	
DR Biotechnol. Appl. Biochem. 19:141-153(1994).	
DR HSSP: P07987; IC82.	
SO SEQUENCE 38 AA; 3896 MW; D3E4C8BF8A834E17 CRC64;	

Query Match	37.5%; Score 3; DB 3; Length 38;
Best Local Similarity	100.0%; Pred. No. 6; 7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY 5 syxg 8	
DB 4 SYXG 7	
RESULT 2	
09GVM9	PRELIMINARY; PRT; 77 AA.
ID 09GVM9;	
AC 09GVM9;	
DT 01-MAR-2001 (TREMBLrel. 16, Created)	

```
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE HYPOTHEITICAL 8.1 KDA PROTEIN (FRAGMENT).
OS Leishmania major.
OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FRIEDLIN;
RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,
RA Oliver K.;
RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AL390114; CAC02854.1; -.
KW Hypothetical protein.
FT NON_TER 1
FT SEQUENCE 77 AA; 8080 MW; CDBCF973C60CA844 CRC64;

Query Match
Best Local Similarity 37.5%; Score 3; DB 5; Length 77;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 SYXG 8
DB 46 SYXG 49

RESULT 3
ID Q9MZE1 PRELIMINARY; PRT; 91 AA.
AC Q9MZE1;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE GLUTATHIONE PEROXIDASE (FRAGMENT).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Vermette L., Stirois J., Carriere P., Price C., Silversides D.W.,
RA Lussier J.G.;
RT "Cow glutathione peroxidase.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF236854; AAF74263.1; -.
DR InterPro: IPR000889; Glut_peroxidase.
DR Pfam: PF00255; GSHPx; 1.
KW Peroxidase.
FT NON_TER 91
FT SEQUENCE 91 AA; 10018 MW; 194E4AE861F41C74 CRC64;

Query Match
Best Local Similarity 37.5%; Score 3; DB 6; Length 91;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 SYXG 8
DB 71 SYXG 74

RESULT 4
ID Q9I947 PRELIMINARY; PRT; 131 AA.
AC Q9I947;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DR EY43 HOMOLOGUE (FRAGMENT).
```

```
OS Oryzias latipes (Medaka fish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.
OX NCBI_TaxID=8090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HNI; TISSUE=WHOLE EMBRYO;
RA Morita Y., Mitani H., Naruse K.;
RT "Oryzias latipes eya3 homologue partial sequence.";
RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AB032898; BAA97579.1; -.
FT NON_TER 1
FT NON_TER 131
FT SEQUENCE 131 AA; 14225 MW; AA3D41A25564E9C3 CRC64;

Query Match
Best Local Similarity 37.5%; Score 3; DB 13; Length 131;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 SYXG 8
DB 67 SYXG 70

RESULT 5
ID Q9GWB2 PRELIMINARY; PRT; 136 AA.
AC Q9GWB2;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE PROBABLE SINGLE STRAND-SPECIFIC NUCLEASE (FRAGMENT).
GN IM12.981.
OS Leishmania major.
OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FRIEDLIN;
RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,
RA Oliver K.;
RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AL390114; CAC02622.1; -.
FT NON_TER 1
FT NON_TER 136
FT SEQUENCE 136 AA; 14910 MW; 018F61A2600F9BC CRC64;

Query Match
Best Local Similarity 37.5%; Score 3; DB 5; Length 136;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 SYXG 8
DB 52 SYXG 55

RESULT 6
ID Q27953 PRELIMINARY; PRT; 141 AA.
AC Q27953;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE B-CASEIN (FRAGMENT).
OS Balanoptera physalus (Finback whale) (Common rosqnal).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
OC Balanopteridae; Balanoptera.
OX NCBI_TaxID=9770;
```

```

RN
[1] 4
SEQUENCE FROM N.A.
RA Galesy J., Hayashi C., Cronin M., Arcander P.;
RA MOL. Biol. Evol. 0:0-0(0).
DR EMBL: U53900; AAB08405.1; -.
DR InterPro: IPR001589; Casein.
DR Pfam: PF00363; caseins; 1.
FT NON_TER 1
FT NON_TER 141 141
SO SEQUENCE 141 AA; 15822 MW; 7C3DEEE320034513 CRC64;

Query Match 37.5%; Score 3; DB 6; Length 141;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
|111|
Db 13 SYXG 16

RESULT 7
ID 09NZ74 PRELIMINARY; PRT; 226 AA.
AC 09NZ74;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE EXTRACELLULAR GLUTATHIONE PEROXIDASE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Comhair S.A.A., Thomassen M.J., Erzurum S.C.;
RT "Differential induction of nitric oxide synthase 2 and extracellular
RT glutathione peroxidase in airways of healthy individuals exposed to
RT 100% O2 or cigarette smoke.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF217787; AAF43005.1; -.
DR InterPro: IPR000889; Glut_peroxidase.
DR Pfam: PF00255; GSHpx; 1.
DR PRINTS: PRO1011; GLUTPROXOASE.
DR PROSITE: PS00763; GLUTATHIONE_PEROXID_2; 1.
KW Peroxidase.
SQ SEQUENCE 226 AA; 25512 MW; 079B970F7C0651A4 CRC64;

Query Match 37.5%; Score 3; DB 4; Length 226;
Best Local Similarity 100.0%; Pred. No. 2.8e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
|111|
Db 71 SYXG 74

RESULT 8
ID 09GYV3 PRELIMINARY; PRT; 485 AA.
AC 09GYV3;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE VARIANT SURFACE GLYCOPROTEIN.
OS Trypanosoma brucei.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID:5691;
RN [1]
RP SEQUENCE FROM N.A.
RA Wickstead B., Ersfeld K., Gull K.;
```

```

RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF294807; AAG03079.1; -.
DR InterPro: IPR001812; Trypan_glycop.
DR Pfam: PF00913; Trypan_glycop; 1.
SQ SEQUENCE 485 AA; 52687 MW; 9D5C330AF1167836 CRC64;

Query Match 37.5%; Score 3; DB 5; Length 485;
Best Local Similarity 100.0%; Pred. No. 5.1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
|111|
Db 326 SYXG 329

RESULT 9
ID 09BB52 PRELIMINARY; PRT; 707 AA.
AC 09BB52;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE NADH DEHYDROGENASE F (FRAGMENT).
GN NDHF.
OS Schoenoecephalum cucullatum.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Commelinidae; Inceratae; sedis;
OC Rapateaceae; Schoenoecephalum.
OX NCBI_TaxID=142441;
RN [1]
RP SEQUENCE FROM N.A.
RA Givnish T.J., Evans T.M., Zjhra M.L., Patterson T.B., Berry P.E.,
RA Systma K.J.;
RT "Molecular evolution, adaptive radiation, and geographic
RT diversification in the amphiatlantic family Rapateaceae: evidence from
RT ndhF sequences and morphology.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF207634; AAK21842.1; -.
KW Chloroplast.
FT NON_TER 1
FT NON_TER 707 707
SQ SEQUENCE 707 AA; 79703 MW; D02C25580E2CB659 CRC64;

Query Match 37.5%; Score 3; DB 8; Length 707;
Best Local Similarity 100.0%; Pred. No. 6.9e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
|111|
Db 523 SYXG 526

RESULT 10
ID 015898 PRELIMINARY; PRT; 8 AA.
AC 015898;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMBlrel. 07, Last annotation update)
DE (CLONE XP6A11B) (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Lee C.C., Yazdani A., Wehnert M., Bailey J., Couch L., Xiong M.,
RA Coolbaugh M.I., Chinault C.A., Baldini A., Lindsay E.A., Zhao Z.Y.,
RA Caskey C.T.H.;
```

RL Hum. Mol. Genet. 0:0-0(0).
DR EMBL: J32078; AAA73888.1; -.
FT NON_TER 1 1
FT NON_TER 8 8
SQ SEQUENCE 8 AA; 938 MW; 34A415B0477B45B8 CRC64;

Query Match 25.0%; Score 2; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
11
DB 2 SY 3

RESULT 11
ID 034909 PRELIMINARY; PRT; 8 AA.

DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE CYTOCHROME B (FRAGMENT).
OS Locusta migratoria (Migratory locust).
OC Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caellifera;
OC Acridomorpha; Acridoidea; Acrididae; Locusta.
OX NCBI_TaxID=7004;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88223478; PubMed=2836084;
RA McCracken A., Uhlenbusch T., Gellissen G.;
RT "Structure of the cloned Locusta migratoria mitochondrial genome:
restriction mapping and sequence of its ND-1 (URF-1) gene.";
RL Curr. Genet. 11:625-630(1987).
DR EMBL: X05286; CAA28905.1; -.
KW Mitochondrion.
FT NON_TER 1 1
SQ SEQUENCE 8 AA; 1019 MW; F8E33723304B45B6 CRC64;

Query Match 25.0%; Score 2; DB 8; Length 8;
Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
11
DB 2 SY 3

RESULT 12
ID 03S824 PRELIMINARY; PRT; 8 AA.

DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE PHOTOSYSTEM I REACTION CENTER SUBUNIT IV, PSI-E.
OS Spinacia oleracea (Spinach).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.
OX NCBI_TaxID=3562;
RN [1]
RP SEQUENCE.

RX MEDLINE=92249324; PubMed=1374333;
RA Lagoute B., Vallon O.;
RT "Purification and membrane topology of PSI-D and PSI-E, two subunits
of the photosystem I reaction center.";
RL Eur. J. Biochem. 205:1175-1185(1992).
SQ SEQUENCE 8 AA; 1082 MW; 2145BBI324069044 CRC64;

Query Match 25.0%; Score 2; DB 10; Length 8;
Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
11
DB 7 SY 8

RESULT 13
ID 062721 PRELIMINARY; PRT; 8 AA.

DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE PROHIBITIN (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER;
RX MEDLINE=95331633; PubMed=7607556;
RA Altus M.S., Wood C.M., Stewart D.A., Roskams A.L., Friedman V.,
RA Henderson T., Owens G.A., Danner D.B., Jupe E.R., Dell'Orco R.T.,
RA McClung J.K.;
RT "Regions of evolutionary conservation between the rat and human
prohibitin-encoding genes.";
RL Gene 158:291-294(1995).
DR EMBL: U17178; AAA86692.1; -.
FT NON_TER 8 8
SQ SEQUENCE 8 AA; 1150 MW; EFD3237B05A41376 CRC64;

Query Match 25.0%; Score 2; DB 11; Length 8;
Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 wr 2
11
DB 2 WR 3

RESULT 14
ID 09BYF9 PRELIMINARY; PRT; 9 AA.

DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CYTOKERATIN 19 (FRAGMENT).
CN K19.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kataya M., Kaneko S., Ohno H., Inamura K., Kobayashi K.;
RT "Promoter activity and protein binding sites of the regulatory
sequences of the human cytokeratin 19 gene.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB045973; BAB40770.1; -.
KW Keratin.
FT NON_TER 9 9
SQ SEQUENCE 9 AA; 1122 MW; 9E9FC41B45A45A1 CRC64;

Query Match 25.0%; Score 2; DB 4; Length 9;

Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
11
Db 3 sy 4

RESULT 15
Q28093
ID Q28093 PRELIMINARY; PRT: 9 AA.

AC Q28093;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE CYTOKERATIN IV GENE UPSTREAM REGION (FRAGMENT).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89231609; PubMed=2469572;
RT Blessing M., Jorcane J.L., Franke W.W.;
RT "Enhancer elements directing cell-type-specific expression of
RT cytokerin genes and changes of the epithelial cytoskeleton by
RT transfections of hybrid cytokerin genes.";
RL EMBO J. 8:117-126(1989).
KM EMBL: X14478; CAA32640.1; -.
DN Keratin: Intermediate filament.
FT NON_TER 9
SQ SEQUENCE 9 AA; 1044 MW; 819A22D1B5B32B45 CRC64;

Query Match 25.0%; Score 2; DB 6; Length 9;
Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
11
Db 2 sy 3

RESULT 16
Q9QWTO
ID Q9QWTO PRELIMINARY; PRT: 9 AA.

AC Q9QWTO;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE PROTEINASE 3 (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathii; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SV129 D3;
RA Lutz P.G., Moog-Lutz C., Houzel-Charavel A., Cayre Y.E.;
RA Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL: AJ007030; CAA07429.1; -.
FT NON_TER 9
SQ SEQUENCE 9 AA; 937 MW; C91E75A7B45B87D CRC64;

Query Match 25.0%; Score 2; DB 11; Length 9;
Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
11
Db 4 sy 5

RESULT 17
Q91A14
ID Q91A14 PRELIMINARY; PRT: 9 AA.

AC Q91A14;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE FIBROBLAST GROWTH FACTOR HOMOLOGOUS FACTOR 2 ISOFORM 1U (FRAGMENT).
GN FGF-2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20112823; PubMed=10644718;
RA Munoz-Sanjuan I., Smallwood P.M., Nathans J.;
RT "Isoform Diversity among Fibroblast Growth Factor Homologous Factors
RT Is Generated by Alternative Promoter Usage and Differential
RT Splicing.";
RL J. Biol. Chem. 275:2589-2597(2000).
DR EMBL: AF199609; AAF31396.1; -.
FT NON_TER 9
SQ SEQUENCE 9 AA; 1068 MW; DF9245B32407272D CRC64;

Query Match 25.0%; Score 2; DB 13; Length 9;
Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
11
Db 7 sy 8

RESULT 18
Q9UC46
ID Q9UC46 PRELIMINARY; PRT: 11 AA.

AC Q9UC46;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE NEUTROPHIL INHIBITOR PEPTIDE, NIP-POLYMORPHONUCLEAR NEUTROPHIL
DE INHIBITOR PEPTIDE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=96326114; PubMed=8703476;
RA Cooper J.A.Jr., Culbreth R.R.;
RA "Characterization of a neutrophil inhibitor peptide harvested from
RT human bronchial lavage: homology to influenza A nucleoprotein.";
RL Am. J. Respir. Cell Mol. Biol. 15:207-215(1996).
SQ SEQUENCE 11 AA; 1262 MW; 951A1C3279C9DB45 CRC64;

Query Match 25.0%; Score 2; DB 4; Length 11;
Best Local Similarity 100.0%; Pred. No. 5.1e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
11
Db 4 sy 5

RESULT 19
Q77876
ID Q77876 PRELIMINARY; PRT: 11 AA.

AC 077876;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE MHC CLASS II B LOCUS 1 (FRAGMENT).
 OC Oreochromis niloticus (Nile tilapia) (Tilapia nilotica).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 CC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Labroidae;
 CC Cichlidae; Oreochromis.
 OX NCBI_TaxID=8128;
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98315113; PubMed=9649539;
 RA Malaga-Trillo E., McAndrew B., Vincek V., Zaleska-Rutczynska Z.,
 RA Suelmann H., Figueroa F., Klein J.;
 RT "Linkage relationships and haplotype polymorphism among cichlid Mhc
 class II B loci.";
 RL Genetics 149:1527-1547(1998).
 DR EMBL; AF049985; AAC41324.1; -.
 KW MHC.
 FT NON_TER 1 1
 FT NON_TER 11 11
 SQ SEQUENCE 11 AA; 1178 MW; 9AC131FAB2D2DB45 CRC64;

Query Match 25.0%; Score 2; DB 7; Length 11;
 Best Local Similarity 100.0%; Pred. No. 5.1e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
 II
 DB 4 sy 5

RESULT 20
 077883
 ID 077883 PRELIMINARY; PRT; 11 AA.

DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE MHC CLASS II B LOCUS 4 (FRAGMENT).
 OC Oreochromis niloticus (Nile tilapia) (Tilapia nilotica).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 CC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Labroidae;
 CC Cichlidae; Oreochromis.
 OX NCBI_TaxID=8128;
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98315113; PubMed=9649539;
 RA Malaga-Trillo E., McAndrew B., Vincek V., Zaleska-Rutczynska Z.,
 RA Suelmann H., Figueroa F., Klein J.;
 RT "Linkage relationships and haplotype polymorphism among cichlid Mhc
 class II B loci.";
 RL Genetics 149:1527-1547(1998).
 DR EMBL; AF049992; AAC4131.1; -.
 KW MHC.
 FT NON_TER 1 1
 FT NON_TER 11 11
 SQ SEQUENCE 11 AA; 1288 MW; 87F42A0FB2D5AB45 CRC64;

Query Match 25.0%; Score 2; DB 7; Length 11;
 Best Local Similarity 100.0%; Pred. No. 5.1e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
 II
 DB 4 sy 5

RESULT 21
 077886
 ID 077886 PRELIMINARY; PRT; 11 AA.
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE MHC CLASS II B LOCUS 4 (FRAGMENT).
 OC Oreochromis niloticus (Nile tilapia) (Tilapia nilotica).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 CC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Labroidae;
 CC Cichlidae; Oreochromis.
 OX NCBI_TaxID=8128;
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98315113; PubMed=9649539;
 RA Malaga-Trillo E., McAndrew B., Vincek V., Zaleska-Rutczynska Z.,
 RA Suelmann H., Figueroa F., Klein J.;
 RT "Linkage relationships and haplotype polymorphism among cichlid Mhc
 class II B loci.";
 RL Genetics 149:1527-1547(1998).
 DR EMBL; AF049995; AAC41334.1; -.
 KW MHC.
 FT NON_TER 1 1
 FT NON_TER 11 11
 SQ SEQUENCE 11 AA; 1343 MW; 87F42D9F52D41B45 CRC64;

Query Match 25.0%; Score 2; DB 7; Length 11;
 Best Local Similarity 100.0%; Pred. No. 5.1e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
 II
 DB 4 sy 5

RESULT 22
 077912
 ID 077912 PRELIMINARY; PRT; 11 AA.
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE MHC CLASS II B LOCUS 4 (FRAGMENT).
 OC Oreochromis niloticus (Nile tilapia) (Tilapia nilotica).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 CC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Labroidae;
 CC Cichlidae; Oreochromis.
 OX NCBI_TaxID=8128;
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98315113; PubMed=9649539;
 RA Malaga-Trillo E., McAndrew B., Vincek V., Zaleska-Rutczynska Z.,
 RA Suelmann H., Figueroa F., Klein J.;
 RT "Linkage relationships and haplotype polymorphism among cichlid Mhc
 class II B loci.";
 RL Genetics 149:1527-1547(1998).
 DR EMBL; AF050023; AAC41362.1; -.
 KW MHC.
 FT NON_TER 1 1
 FT NON_TER 11 11
 SQ SEQUENCE 11 AA; 1288 MW; 87F42A0FB2D5AB45 CRC64;

Query Match 25.0%; Score 2; DB 7; Length 11;
 Best Local Similarity 100.0%; Pred. No. 5.1e+03;
 Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
 II

```

Db          4  5
RESULT 23
ID 032704      PRELIMINARY;      PRT:      11 AA.
AC 032704:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE NDHE GENE PRODUCT (12 AA) (FRAGMENT).
GN NDHE.
OS Nicotiana tabacum (Common tobacco).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxId=4097;
RN
RP
RP SEQUENCE FROM N.A.
RC STRAIN=BRIGHT YELLOW 4; TISSUE=LEAVE;
MEDLINE=88210537; PubMed=3329576;
RA Hayashida N., Matsubayashi T., Shinozaki K., Suglura M., Inoue K.,
RA Hiyaama T.;
RT "The gene for the 9 kd polypeptide, a possible apoprotein for the
RT iron-sulfur centers A and B of the photosystem I complex, in tobacco
RT chloroplast DNA."
RL Curr. Genet. 12:247-250(1987).
DR EMBL; X05881; CAA29303.1; -.
KM Chloroplast.
FT NON_TER
SO SEQUENCE 11 AA; 1338 MW; 008165EE304776CB CRC64;

Query Match          25.0%; Score 2; DB 8; Length 11;
Best Local Similarity 100.0%; Pred. No. 5.1e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
   11
Db 3 SY 4

RESULT 24
ID 09G607      PRELIMINARY;      PRT:      11 AA.
AC 09G607:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CYTOCHROME C OXIDASE SUBUNIT I (FRAGMENT).
GN COI.
OS Aphanolotis fusca.
OS Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Iguania; Acrodonta; Agamidae; Draconinae;
OC Aphanolotis.
OX NCBI_TaxId=89036;
RN
RP
RP SEQUENCE FROM N.A.
RC Macey J.R., Schulte J.A. II, Larson A.;
RT "Evolution and information content of the mitochondrial genomic
RT structural features illustrated with acrodont lizards."
RL Syst. Biol. 49:257-277(2000).
RN
RP
RP SEQUENCE FROM N.A.
RC Macey J.R., Schulte J.A. II, Larson A., Ananjeva N.B., Wang Y.,
RA Bethiyagoda R., Rastegar-Pouyani N., Papenfuss T.J.;
RT "Evaluating Trans-Tethys migration: An example using Acrodont lizard
RT phylogenetics."
RL Syst. Biol. 49:233-256(2000).
DR EMBL; AF128495; AAG00722.1; -.
KM Mitochondrion.
FT NON_TER
SO SEQUENCE 11 AA; 1341 MW; 538E371E33640DD7 CRC64;

```

```

FT NON_TER 11 11
SQ SEQUENCE 11 AA; 1369 MW; A68E371E336411A6 CRC64;

Query Match          25.0%; Score 2; DB 8; Length 11;
Best Local Similarity 100.0%; Pred. No. 5.1e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
   11
Db 2 SY 3

RESULT 25
ID 09G359      PRELIMINARY;      PRT:      11 AA.
AC 09G359:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CYTOCHROME C OXIDASE SUBUNIT I (FRAGMENT).
GN COI.
OS Japalura flaviceps.
OS Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Iguania; Acrodonta; Agamidae; Draconinae;
OC Japalura.
OX NCBI_TaxId=52218;
RN
RP
RP SEQUENCE FROM N.A.
RC MEDLINE=97153820; PubMed=9000751;
RA Macey J.R., Larson A., Ananjeva N.B., Papenfuss T.J.;
RT "Replication slippage may cause parallel evolution in the secondary
RT structures of mitochondrial transfer RNAs."
RL Mol. Biol. Evol. 14:30-39(1997).
RN
RP
RP SEQUENCE FROM N.A.
RC Macey J.R., Schulte J.A. II, Larson A.;
RT "Evolution and information content of the mitochondrial genomic
RT structural features illustrated with acrodont lizards."
RL Syst. Biol. 49:257-277(2000).
RN
RP
RP SEQUENCE FROM N.A.
RC Macey J.R., Schulte J.A. II, Larson A., Ananjeva N.B., Wang Y.,
RA Bethiyagoda R., Rastegar-Pouyani N., Papenfuss T.J.;
RT "Evaluating Trans-Tethys migration: An example using Acrodont lizard
RT phylogenetics."
RL Syst. Biol. 49:233-256(2000).
DR EMBL; AF128500; AAG00737.1; -.
KM Mitochondrion.
FT NON_TER
SO SEQUENCE 11 AA; 1341 MW; 538E371E33640DD7 CRC64;

Query Match          25.0%; Score 2; DB 8; Length 11;
Best Local Similarity 100.0%; Pred. No. 5.1e+03;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 sy 6
   11
Db 2 SY 3

Search completed: January 14, 2002, 08:07:34
Job time: 764 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 07:56:31 ; Search time 81.39 Seconds
(without alignments)
7.281 Million cell updates/sec

Title: 09-185908-1E
Sequence: 1 wrxxxxyxg 8

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 50 summaries

Database :

A_Geneseq_1101: *
1: /SIDs2/gcgdata/geneseq/geneseqp/AA1980.DAT: *
2: /SIDs2/gcgdata/geneseq/geneseqp/AA1981.DAT: *
3: /SIDs2/gcgdata/geneseq/geneseqp/AA1982.DAT: *
4: /SIDs2/gcgdata/geneseq/geneseqp/AA1983.DAT: *
5: /SIDs2/gcgdata/geneseq/geneseqp/AA1984.DAT: *
6: /SIDs2/gcgdata/geneseq/geneseqp/AA1985.DAT: *
7: /SIDs2/gcgdata/geneseq/geneseqp/AA1986.DAT: *
8: /SIDs2/gcgdata/geneseq/geneseqp/AA1987.DAT: *
9: /SIDs2/gcgdata/geneseq/geneseqp/AA1988.DAT: *
10: /SIDs2/gcgdata/geneseq/geneseqp/AA1989.DAT: *
11: /SIDs2/gcgdata/geneseq/geneseqp/AA1990.DAT: *
12: /SIDs2/gcgdata/geneseq/geneseqp/AA1991.DAT: *
13: /SIDs2/gcgdata/geneseq/geneseqp/AA1992.DAT: *
14: /SIDs2/gcgdata/geneseq/geneseqp/AA1993.DAT: *
15: /SIDs2/gcgdata/geneseq/geneseqp/AA1994.DAT: *
16: /SIDs2/gcgdata/geneseq/geneseqp/AA1995.DAT: *
17: /SIDs2/gcgdata/geneseq/geneseqp/AA1996.DAT: *
18: /SIDs2/gcgdata/geneseq/geneseqp/AA1997.DAT: *
19: /SIDs2/gcgdata/geneseq/geneseqp/AA1998.DAT: *
20: /SIDs2/gcgdata/geneseq/geneseqp/AA1999.DAT: *
21: /SIDs2/gcgdata/geneseq/geneseqp/AA2000.DAT: *
22: /SIDs2/gcgdata/geneseq/geneseqp/AA2001.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3	37.5	7	19	AAW65247	Peptide containing
2	3	37.5	24	20	AAW74457	Fibroblast growth
3	3	37.5	24	21	AAV90893	Fibroblast growth
4	3	37.5	34	22	AAW15839	Peptide #2275 enco
5	3	37.5	34	22	AAW28350	Peptide #2387 enco
6	3	37.5	34	22	AAW03577	Peptide #2259 enco
7	3	37.5	53	11	AAW08007	Modified murine ep
8	3	37.5	72	22	AAW77776	Human colon cancer
9	3	37.5	90	21	AAW28199	Arabidopsis thalia
10	3	37.5	96	21	AAW28198	Arabidopsis thalia
11	3	37.5	104	14	AAW38596	Human lambda light

12	3	37.5	104	19	AAW58496	Human lambda light
13	3	37.5	105	21	AAW02234	Human secreted pro
14	3	37.5	129	22	AAW57533	Human colon cancer
15	3	37.5	158	22	AAW75907	Human colon cancer
16	3	37.5	164	21	AAW28197	Arabidopsis thalia
17	3	37.5	226	16	AAW72680	Human plasma gluta
18	3	37.5	226	16	AAW22828	Human plasma gluta
19	3	37.5	230	18	AAW08430	Human plasma gluta
20	3	37.5	280	21	AAW08728	Rapamycin-depend
21	3	37.5	329	21	AAW08730	Amilo acid sequenc
22	3	37.5	533	15	AAW44893	Diphtheria toxin (
23	3	37.5	684	17	AAW17581	Thermoanaerobacter
24	3	37.5	684	17	AAW17585	Thermoanaerobacter
25	3	37.5	684	17	AAW17589	Thermoanaerobacter
26	3	37.5	685	17	AAW17582	Thermoanaerobacter
27	3	37.5	685	17	AAW17586	Thermoanaerobacter
28	3	37.5	685	17	AAW17590	Thermoanaerobacter
29	2	25.0	3	15	AAW48523	Lactoferrin deriva
30	2	25.0	3	16	AAW84691	Bovine lactoferrin
31	2	25.0	3	17	AAW98547	Peptide for anti-u
32	2	25.0	3	17	AAW90601	Lactoferrin deriva
33	2	25.0	3	19	AAW56219	Anti-inflammatory
34	2	25.0	3	19	AAW41282	Apoptosis inducer
35	2	25.0	3	22	AAW92005	PMR neuropeptide
36	2	25.0	4	2	AAW10154	Sequence of cypto
37	2	25.0	4	2	AAW10370	Genetic enkephalin
38	2	25.0	4	2	AAW10372	Enkephalin-like an
39	2	25.0	4	2	AAW10373	Enkephalin-like an
40	2	25.0	4	2	AAW10375	Enkephalin-like an
41	2	25.0	4	2	AAW10386	Generic analgesic
42	2	25.0	4	2	AAW10599	N-adamantane tetra
43	2	25.0	4	2	AAW10433	Analgesic tetrapep
44	2	25.0	4	2	AAW10401	Agonist peptide.
45	2	25.0	4	2	AAW10620	Analgesic tetrapep
46	2	25.0	4	2	AAW10625	Analgesic tetrapep
47	2	25.0	4	3	AAW20208	Analgesic and neur
48	2	25.0	4	3	AAW20210	Analgesic and neur
49	2	25.0	4	5	AAW40339	Sequence of enkeph
50	2	25.0	4	7	AAW61359	Sequence of peptid

ALIGNMENTS

RESULT 1	
AAW65247	standard; peptide; 7 AA.
XX	
AC	AAW65247;
XX	
DT	02-OCT-1998 (first entry)
XX	
DE	Peptide containing 6-peptidylamino-1-naphthylenesulphonamide moiety.
XX	
XX	Aminonaphthylenesulphonamide; activated protein C; blood coagulation;
KW	APC; substrate; peptidomimetic.
KM	
XX	
OS	Synthetic.
XX	
FH	Key
FT	Modified-site
FT	Location/Qualifiers
FT	1 /note= "N-terminally modified by t-butoxycarbonyl (BOC)"
FT	4 /note= "Homocysteine"
FT	Modified-site
FT	5 /note= "Aminonaphthylenesulphonamide moiety"
FT	7 /note= "Glu(Obz)"
XX	
PN	W09822125-A1.
XX	
PD	28-MAY-1998.
XX	

```
PF 18-NOV-1997; 97WO-US21075.
XX
PR 19-NOV-1996; 96US-0031359.
XX
PA (BUTE/) BUTENAS S.
XX (MANN/) MANN K G.
PI Butenas S, Mann KG;
XX WPI: 1998-312167/27.
XX
XX Peptidyl-amino-1-naphthalene-sulphonamide compounds - are
PT peptidomimetics of natural substrates for activated protein C,
PR useful as promoters of the blood coagulation process
XX
XX Claim 6; Page 38; 50pp; English.
XX
XX The invention relates to peptidomimetics containing 6-peptidylamino-1-
CC naphthalenesulphonamide moieties. They are peptidomimetics of natural
CC substrates for activated protein C (APC), whose substrates are promoters
CC of the blood coagulation process. The present sequence represents a
CC specifically claimed peptidomimetic.
XX
XX Sequence 7 AA;
SQ

Query Match 37.5%; Score 3; DB 19; Length 7;
Best Local Similarity 100.0%; Pred. No. 4.3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 wxxs 5
   |||||
DB 2 wxxs 6

RESULT 2
AAM74457
ID AAM74457 standard; peptide; 24 AA.
XX
AC AAM74457;
XX
XX 12-MAY-1999 (first entry)
DE Fibroblast growth factor 3 receptor mutational cassette.
XX
XX Agonist identification: orphan receptor; constitutively active OR;
KW Graves' disease; thyroid adenoma; hypertension; cardiomyopathy;
KW schizophrenia; Kaposi's sarcoma; fibroblast growth factor receptor;
KW adenylate cyclase constitutive activator; thyrotropin receptor;
KW thyrotropin stimulating hormone; beta-adrenergic receptor.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH Misc-difference 9 /label= Glu, Asp, Gln, His, Lys
FT
PN WO9846995-A1.
XX
XX 22-OCT-1998.
XX
XX 14-APR-1998; 98WO-US07496.
XX
XX 14-APR-1997; 97US-0839449.
XX
XX (BEHA/) BEHAN D P.
PA (CHAL/) CHALMERS D T.
XX
XX Behan DP, Chalmers DT;
PI
XX WPI: 1999-105468/09.
DR
XX
XX Identifying agonists of orphan receptors from their effect on the
PT
```

```
PT constitutively active receptor - particularly therapeutically active
PT inverse agonists at G protein coupled receptors, without requiring
PT knowledge of endogenous ligand or receptor function
XX
XX Disclosure; Page 47; 114pp; English.
XX
XX This sequence is a fragment of a fibroblast growth factor 3 receptor
CC mutational cassette. The invention relates to a method for the
CC identification of candidate compounds as agonists, including inverse or
CC partial, of an orphan receptor (OR), which comprises: (i) applying test
CC compound to constitutively active OR; and (ii) measuring its effect on
CC OR. The method is particularly used to identify inverse agonists of
CC G protein-coupled OR, i.e. potential therapeutic agents for treating
CC conditions in which constitutively active OR are implicated (e.g. Graves'
CC disease, thyroid adenoma, hypertension, cardiomyopathy, schizophrenia,
CC major depression, Kaposi's sarcoma and many others tabulated). It is
CC based on identification of agents that reduce receptor activation, rather
CC than compounds that antagonise the normal ligand. Once identified,
CC (inverse) agonists can be used to study OR function. The method does
CC not require knowledge of the endogenous receptor ligand or receptor
CC function, and identifies directly compounds that inhibit the activated
CC receptor, i.e. able to block both ligand-dependent and -independent
CC activation, rather than only the ligand-dependent process, as is the
CC case with compounds identified by ligand-dependent assays. It should
CC accelerate drug discovery at a wide range of OR and since activated
CC receptors have a greater response to the agents, potential drugs are more
CC likely to be detected.
XX
XX Sequence 24 AA;
SQ
```

```
Query Match 37.5%; Score 3; DB 20; Length 24;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 5 syxx 8
   |||||
DB 7 syxx 10
```

```
RESULT 3
AA90893
ID AA90893 standard; peptide; 24 AA.
XX
AC AA90893;
XX
XX 30-AUG-2000 (first entry)
DE Fibroblast growth factor 3 receptor mutational cassette #1.
XX
XX Identification; modulator; cell surface membrane receptor; treatment;
KW orphan receptor; antithyroid; antidiabetic; neuroleptic; antidepressant;
KW cytosstatic; G protein-coupled receptor agonist.
XX
XX Synthetic.
OS
XX
XX Key Location/Qualifiers
FH Misc-difference 9 /label= Glu, Asp, Gln, His, Lys
FT
PN WO200021987-A2.
XX
XX 20-APR-2000.
XX
XX 12-OCT-1999; 99WO-US23935.
XX
XX 13-OCT-1998; 98US-0170496.
XX
XX (AREN-) ARENA PHARM INC.
PA
XX Behan DP, Chalmers DT;
PI
XX WPI: 2000-317935/27.
DR
XX
XX
```

XX Identifying compounds with inverse agonist activity to orphan receptors
PT useful for treating e.g. Graves' disease, and schizophrenia, involves
PT contacting candidate compounds with constitutively activated receptors
PT
XX
PS Disclosure; Page 47; 110pp; English.
XX
CC The present invention describes a method for directly identifying a
CC compound having inverse agonist activity, partial agonist activity or
CC agonist activity to a constitutively active orphan receptor (ORR). The
CC method comprises determining the efficacy of the compound by contacting
CC it with the ORR. A compound identified by the above method having
CC inverse agonist activity to ORR is useful for the treatment of diseases
CC characterised by constitutive activation of the receptor e.g. Graves'
CC disease, male precocious puberty, Jansen's disease, retinitis pigmentosa,
CC hypoparathyroidism, neuropsychiatric diseases, schizophrenia, major
CC depression, and cancerous growth in Kaposi's sarcoma. The method can
CC identify (1) directly without prior knowledge or use of receptor ligands
CC and is useful for accelerating drug discovery at a broad range of ORR.
CC The present sequence represents a fibroblast growth factor 3 receptor
CC mutational cassette, which is used in the exemplification of
CC the present invention.

SO Sequence 24 AA:

Query Match Best Local Similarity 37.5%; Score 3; DB 21; Length 24;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxx 8
||||
DB 7 syxx 10

RESULT 4

AA015839 AAM15839 standard; Protein; 34 AA.

AC AAM15839;

DT 12-OCT-2001 (first entry)

DE Peptide #2273 encoded by probe for measuring cervical gene expression.

KW Probe: human; microarray; gene expression; cervical epithelial cell;
cervical cancer.

OS Homo sapiens.

PN WO200157278-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00670.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

WIPI: 2001-488901/53.

Human genome-derived single exon nucleic acid probes useful for
analyzing gene expression in human cervical epithelial cells -

XX Claim 27; SEQ ID No 20665; 487pp; English.

PS The present invention relates to human single exon nucleic acid probes
XX (SNP: see A010068-A018459). The present sequence is a peptide encoded
CC by one such probe. The SNPs are derived from human HeLa cells. The SNPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published.pct_sequences.

SO Sequence 34 AA:

Query Match Best Local Similarity 37.5%; Score 3; DB 22; Length 34;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxx 8
||||
DB 18 syxx 21

RESULT 5

AA028350 AAM28350 standard; Protein; 34 AA.

AC AAM28350;

DT 18-OCT-2001 (first entry)

DE Peptide #2387 encoded by probe for measuring placental gene expression.

KW Probe: microarray; human; placenta; antenatal diagnosis;
genetic disorder.

OS Homo sapiens.

PN WO200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

WIPI: 2001-48897/53.

Human genome-derived single exon nucleic acid probes useful for
analyzing gene expression in human placenta -

Claim 27; SEQ ID No 28619; 654pp; English.

The present invention relates to single exon nucleic acid probes (SNP:
see A013115-A0157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.

XX Sequence 34 AA;

Query Match 37.5%; Score 3; DB 22; Length 34;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
| | | |
Db 18 syxg 21

RESULT 6
AAM03577
ID AAM03577 standard; Protein: 34 AA.

XX AAM03577;
XX
DT 09-OCT-2001 (first entry)

XX Peptide #2259 encoded by probe for measuring breast gene expression.

XX probe; human; breast disease; breast cancer; development disorder;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX Homo sapiens.

XX WO200157270-A2.

XX 09-AUG-2001.

XX 29-JAN-2001; 2001WO-US00661.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-476286/51.

XX Novel single exon nucleic acid probe used to measuring gene expression

XX in a human breast -

XX Claim 27: SEQ ID NO 12317; 322pp; English.

XX The present invention relates to novel single exon nucleic acid probes
CC (see AAI00010-AA110067). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for measuring human gene expression in
CC a human breast sample, where the probe hybridises at high stringency to a
CC nucleic acid expressed in the human breast. The probes are useful for
CC predicting, diagnosing, grading, staging, monitoring and prognosing
CC diseases of the human breast, particularly those diseases with polygenic
CC aetiology. The diseases include: breast cancer, disorders of development,
CC inflammatory diseases of the breast, fibrocystic changes, proliferative
CC breast disease and non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 34 AA;

Query Match 37.5%; Score 3; DB 22; Length 34;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
| | | |
Db 18 syxg 21

RESULT 7
AAR08007
ID AAR08007 standard; Protein: 53 AA.

XX AAR08007;

XX 25-FEB-1991 (first entry)

XX Modified murine epidermal growth factor.

XX Modified murine epidermal growth factor; stability; storage;
KW epithelial wounds; gastric acid secretion.

XX Key Location/Qualifiers

XX MISC-difference 11..11 /label= E, N, Q, A, K

XX WO9013570-A.

XX 15-NOV-1990.

XX 09-MAY-1990; 90WO-US02600.

XX 12-MAY-1989; 89US-0351773.

XX (CHIR-) CHIRON CORP.

XX Nascimento CG, Medina-Selby A;

XX WPI; 1990-361427/48.

XX Human epidermal growth factor - is substituted at position 11 for
XX greater stability and improved storage life.

XX Claim 9; Page 25; 32pp; English.

XX The human EGF is used to treat oversecretion of gastric acid or an
CC epithelial wound. EGF is modified to increase its chemical
CC stability. Its storage life is improved without diminishing its
CC biological activity. The proteins may be prepared by traditional
CC chemical or recombinant means.
CC See also AAR08004.

XX Sequence 53 AA;

Query Match 37.5%; Score 3; DB 11; Length 53;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
| | | |
Db 9 syxg 12

RESULT 8

AA677776
ID AA677776 standard; Protein: 72 AA.

XX AA677776;

XX 03-SEP-2001 (first entry)

XX Human colon cancer antigen protein SEQ ID NO:8542.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.

XX OS Homo sapiens.
XX PN WO200122920-A2.
XX PD 05-APR-2001.
XX PF 28-SEP-2000; 2000MO-US26524.
XX PR 29-SEP-1999; 99US-0157137.
XX PR 03-NOV-1999; 99US-0163280.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Ruben SM, Barash SC, Barse CE, Rosen CA;
XX DR MPI; 2001-235357/24.
XX DR N-PSDB; AAH37183.
XX PS Claim 11; Page 9791-9792; 9803pp; English.
XX CC AAH32943 to AAH37195 and AAC73514 to AAC77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAH77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX SO Sequence 72 AA:

Query Match 37.5%; Score 3; DB 22; Length 72;
Best Local Similarity 100.0%; Pred. No. 5.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
|||
DB 54 syxg 57

RESULT 9
ID AAC28199 standard; Protein; 90 AA.
XX AAC28199;
XX AC AAC28199;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 33328.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX KW Arabidopsis thaliana.
XX OS
XX PN EP1033405-A2.

XX XX 06-SEP-2000.
XX PD
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 03-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 23-APR-1999; 99US-0130891.
XX PR 28-APR-1999; 99US-0131449.
XX PR 30-APR-1999; 99US-0132048.
XX PR 04-MAY-1999; 99US-0132407.
XX PR 05-MAY-1999; 99US-0132484.
XX PR 06-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 06-MAY-1999; 99US-0132487.
XX PR 07-MAY-1999; 99US-0132863.
XX PR 11-MAY-1999; 99US-0134256.
XX PR 14-MAY-1999; 99US-0134218.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134221.
XX PR 14-MAY-1999; 99US-0134370.
XX PR 18-MAY-1999; 99US-0134768.
XX PR 19-MAY-1999; 99US-0134941.
XX PR 20-MAY-1999; 99US-0135124.
XX PR 21-MAY-1999; 99US-0135353.
XX PR 24-MAY-1999; 99US-0135629.
XX PR 25-MAY-1999; 99US-0136021.
XX PR 27-MAY-1999; 99US-0136392.
XX PR 28-MAY-1999; 99US-0136782.
XX PR 01-JUN-1999; 99US-0137222.
XX PR 03-JUN-1999; 99US-0137528.
XX PR 04-JUN-1999; 99US-0137502.
XX PR 07-JUN-1999; 99US-0137724.
XX PR 08-JUN-1999; 99US-0138094.
XX PR 10-JUN-1999; 99US-0138540.
XX PR 10-JUN-1999; 99US-0138847.
XX PR 14-JUN-1999; 99US-0139119.
XX PR 16-JUN-1999; 99US-0139452.
XX PR 16-JUN-1999; 99US-0139453.
XX PR 17-JUN-1999; 99US-0139492.
XX PR 18-JUN-1999; 99US-0139454.
XX PR 18-JUN-1999; 99US-0139455.
XX PR 18-JUN-1999; 99US-0139456.
XX PR 18-JUN-1999; 99US-0139457.
XX PR 18-JUN-1999; 99US-0139458.
XX PR 18-JUN-1999; 99US-0139459.
XX PR 18-JUN-1999; 99US-0139460.
XX PR 18-JUN-1999; 99US-0139461.
XX PR 18-JUN-1999; 99US-0139462.
XX PR 18-JUN-1999; 99US-0139463.
XX PR 18-JUN-1999; 99US-0139750.
XX PR 18-JUN-1999; 99US-0139763.
XX PR 21-JUN-1999; 99US-0139817.
XX PR 22-JUN-1999; 99US-0139899.
XX PR 23-JUN-1999; 99US-0140353.
XX PR 23-JUN-1999; 99US-0140354.
XX PR 24-JUN-1999; 99US-0140695.
XX PR 28-JUN-1999; 99US-0140823.
XX PR 29-JUN-1999; 99US-0140991.
XX PR 30-JUN-1999; 99US-0141287.
XX PR 01-JUL-1999; 99US-0141842.
XX PR 01-JUL-1999; 99US-0142154.

[illegible]

PR	22-SEP-1999;	9905-0155139;
PR	23-SEP-1999;	9905-0155466;
PR	24-SEP-1999;	9905-0155565;
PR	28-SEP-1999;	9905-0156458;
PR	29-SEP-1999;	9905-0156596;
PR	04-OCT-1999;	9905-0157117;
PR	05-OCT-1999;	9905-0157753;
PR	06-OCT-1999;	9905-0157865;
PR	07-OCT-1999;	9905-0158029;
PR	08-OCT-1999;	9905-0158232;
PR	12-OCT-1999;	9905-0158369;
PR	13-OCT-1999;	9905-0159293;
PR	13-OCT-1999;	9905-0159294;
PR	14-OCT-1999;	9905-0159329;
PR	14-OCT-1999;	9905-0159330;
PR	14-OCT-1999;	9905-0159331;
PR	14-OCT-1999;	9905-0159637;
PR	14-OCT-1999;	9905-0159638;
PR	18-OCT-1999;	9905-0159584;
PR	21-OCT-1999;	9905-0160741;
PR	21-OCT-1999;	9905-0160767;
PR	21-OCT-1999;	9905-0160981;
PR	21-OCT-1999;	9905-0160768;
PR	21-OCT-1999;	9905-0160770;
PR	21-OCT-1999;	9905-0160814;
PR	21-OCT-1999;	9905-0160815;
PR	22-OCT-1999;	9905-0160980;
PR	22-OCT-1999;	9905-0160981;
PR	22-OCT-1999;	9905-0160989;
PR	25-OCT-1999;	9905-0161404;
PR	25-OCT-1999;	9905-0161405;
PR	25-OCT-1999;	9905-0161406;
PR	26-OCT-1999;	9905-0161359;
PR	26-OCT-1999;	9905-0161360;
PR	26-OCT-1999;	9905-0161361;
PR	28-OCT-1999;	9905-0161920;
PR	28-OCT-1999;	9905-0161922;
PR	28-OCT-1999;	9905-0161993;
PR	29-OCT-1999;	9905-0162142;

Query Match	37.5%	Score 3	DB 21	Length 90
Best Local Similarity	100.0%	Pred. No.	6.5e+02	
Matches	4	Conservative	0	Mismatches 0
		Indels	0	Gaps 0
QY	5 syxg 8			
Db	38 syxg 41			
RESULT	10			
ID	AAG28198			
	AAG28198 standard; Protein; 96 AA.			
XX				
AC	AAG28198;			
XX				
DT	17-OCT-2000 (first entry)			
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 33327.			
XX				
KW	Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.			
XX				
OS	Arabidopsis thaliana.			
XX				
PN	EP1033405-A2.			
XX				
PD	06-SEP-2000.			
XX				
PF	25-FEB-2000; 2000EP-0301439.			
XX				
PR	25-FEB-1999; 990S-0121825.			
PR	05-MAR-1999; 990S-0123180.			

PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0123788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-01300449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136352.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140635.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142340.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145092.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145813.
PR 27-JUL-1999; 99US-0145818.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147503.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151348.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.

PR 06-OCT-1999: 99US-0157865.
PR 07-OCT-1999: 99US-0158029.
PR 08-OCT-1999: 99US-0158232.
PR 12-OCT-1999: 99US-0158369.
PR 13-OCT-1999: 99US-0159293.
PR 13-OCT-1999: 99US-0159294.
PR 13-OCT-1999: 99US-0159295.
PR 14-OCT-1999: 99US-0159329.
PR 14-OCT-1999: 99US-0159330.
PR 14-OCT-1999: 99US-0159331.
PR 14-OCT-1999: 99US-0159637.
PR 14-OCT-1999: 99US-0159638.
PR 18-OCT-1999: 99US-0159584.
PR 21-OCT-1999: 99US-0160741.
PR 21-OCT-1999: 99US-0160767.
PR 21-OCT-1999: 99US-0160768.
PR 21-OCT-1999: 99US-0160770.
PR 21-OCT-1999: 99US-0160814.
PR 21-OCT-1999: 99US-0160815.
PR 22-OCT-1999: 99US-0160980.
PR 22-OCT-1999: 99US-0160981.
PR 22-OCT-1999: 99US-0160989.
PR 25-OCT-1999: 99US-0161404.
PR 25-OCT-1999: 99US-0161405.
PR 25-OCT-1999: 99US-0161406.
PR 26-OCT-1999: 99US-0161359.
PR 26-OCT-1999: 99US-0161360.
PR 26-OCT-1999: 99US-0161361.
PR 28-OCT-1999: 99US-0161920.
PR 28-OCT-1999: 99US-0161992.
PR 28-OCT-1999: 99US-0161993.
PR 29-OCT-1999: 99US-0162142.

Query Match 37.5%; Score 3; DB 21; Length 96;
Best Local Similarity 100.0%; Pred. No. 6.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
||||
Db 44 syxg 47

RESULT 11
AAR38596
ID AAR38596 standard; peptide: 104 AA.

AC AAR38596;

DT 28-OCT-1993 (first entry)

DE Human lambda light chain subgroup 4 (hL4).

XX Antibody; variable domain; light; L; heavy; H; consensus;
KW affinity; antigen; immunogenicity; humanisation; framework.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 1 /note= "residue conserved in less than 50% of the
FT known sequences of hL4"

FT Misc-difference 32 /note= "residue conserved in less than 50% of the
FT known sequences of hL4"

FT Misc-difference 87 /note= "residue conserved in less than 50% of the
FT known sequences of hL4"

FT Misc-difference 89..92 /note= "residues conserved in less than 50% of the
FT known sequences of hL4"

FT W09311794-A.
XX

PD 24-JUN-1993.

XX 14-DEC-1992: 92WO-US10906.

XX 13-DEC-1991: 91US-0808464.

XX (XOMA) XOMA CORP.

XX Fishwild DM, Kohn FR, Little RG, Studnicka GM;

DR WPI; 1993-213827/26.

PT Antibodies prep. used for treatment of auto-immune diseases - by
PT replacement of critical residues to reduce immunogenicity but
PT retain binding affinity, etc.

PS Claim 2; Page 96; 160pp; English.

XX The consensus amino acid sequences for the subgroups of light
CC chains (hK1 - AAR38590, hK3 - NGK, hK2 - GST, hL1 - AAR38591, hL2 -
CC AAR38592, hL3 - AAR38593, hL6 - AAR38594, hK4 - AAR38595, hL4 -
CC AAR38596, and hL5 - AAR38597) and heavy chains (hH3 - AAR38598, hH1 -
CC AAR38599 and hH2 - AAR38600) of human variable domains may be used to
CC prepare, for example, a modified mouse antibody variable domain that
CC retains the affinity of the natural domain for antigen while exhibiting
CC reduced immunogenicity in humans.
CC Unlike other methods of humanisation, which advocate the
CC replacement of entire antibody framework regions with those of human
CC antibodies, this method involves only the introduction of human
CC residues into those positions not critical for antigen binding.
CC This ensures that the binding properties of the modified antibody
CC are not diminished.

XX Sequence 104 AA;

Query Match 37.5%; Score 3; DB 14; Length 104;
Best Local Similarity 100.0%; Pred. No. 7.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
||||
Db 85 syxg 88

RESULT 12

AAW58496
ID AAW58496 standard; protein: 104 AA.

AC AAW58496;

DT 18-AUG-1998 (first entry)

DE Human lambda light chain subgroup 2 consensus sequence hL2.

XX Humanised; human; mouse; CD5; anti-CD5 antibody; immunoglobulin;
KW depletion; cytotoxic; immunoconjugate; fusion protein; psoriasis;
KW autoimmune disease; rheumatoid arthritis; type I diabetes.

XX Synthetic.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 1 /note= "not specified"

FT Misc-difference 32 /note= "not specified"

FT Misc-difference 87 /note= "not specified"

FT Misc-difference 89 /note= "not specified"

FT Misc-difference 90 /note= "not specified"

PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX WPI: 2001-235357/24.
DR N-PSDB: AAH35138.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
PS Claim 11: Page 7974-7975; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated PS,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 129 AA;

Query Match 37.5%; Score 3; DB 22; Length 129;
Best Local Similarity 100.0%; Pred. No. 8.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 syxy 8
DB 117 syxy 120

RESULT 15
AAG75907
ID AAG75907 standard; Protein: 158 AA.
XX
AC AAG75907;
XX
DT 03-SEP-2001 (first entry)
XX
XX Human colon cancer antigen protein SEQ ID NO:6671.
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000MO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
XX
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
XX WPI: 2001-235357/24.
DR N-PSDB: AAH35312.
XX

PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
PS Claim 11: Page 8137-8138; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated PS,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 158 AA;

Query Match 37.5%; Score 3; DB 22; Length 158;
Best Local Similarity 100.0%; Pred. No. 9.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 rxxxy 6
DB 109 rxxxy 113

RESULT 16
AAG28197
ID AAG28197 standard; Protein: 164 AA.
XX
AC AAG28197;
XX
DT 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 33326.
XX
DE Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
XX
PR 05-MAR-1999; 99US-0123180.
XX
PR 09-MAR-1999; 99US-0123548.
XX
PR 23-MAR-1999; 99US-0125788.
XX
PR 25-MAR-1999; 99US-0126264.
XX
PR 29-MAR-1999; 99US-0126785.
XX
PR 01-APR-1999; 99US-0127462.
XX
PR 06-APR-1999; 99US-0128234.
XX
PR 08-APR-1999; 99US-0128714.
XX
PR 16-APR-1999; 99US-0129845.
XX
PR 19-APR-1999; 99US-0130077.
XX
PR 21-APR-1999; 99US-0130449.
XX
PR 23-APR-1999; 99US-0130510.
XX
PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132865.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134376.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141844.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145102.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.

PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 37.5%; Score 3; DB 21; Length 164;
Best Local Similarity 100.0%; Pred. No. 9.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
|||
Db 112 syxg 115

RESULT 17
AAR72680
ID AAR72680 standard; peptide: 226 AA.

AC AAR72680;

DT 01-NOV-1995 (first entry)

DE Human plasma glutathione peroxidase.

KW Human glutathione peroxidase; plasma; antibody; immunoassay; kidney;
selenium deficiency; liver disease; graft rejection; hepatic cancer.

OS Homo sapiens.

FT Key Location/Qualifiers

FT Misc-difference 73 /note= "unknown amino acid"

PN WO9506719-A.

PD 09-MAR-1995.

PF 01-SEP-1994; 94WO-FR01031.

PR 03-SEP-1993; 93FR-0010504.

PA (BIOX-) BIOXYTECH SA.
(BIOX-) BIOXYTECH.

PI Chaudiere J, Lemaingue A, Malette P;

DR WPI; 1995-115435/15.

PT New immunogenic oligo:peptide of human plasma glutathione
peroxidase - and derived antibodies that do not recognise other
PT forms of the enzyme and are useful in immunoassays, e.g. to
PR detect selenium deficiency

PS Disclosure; Fig 1; 33pp; French.

CC The deduced amino acid sequence of the human plasma glutathione
peroxidase as published (J. Biochem., 108, 1990,145-148). Immunogenic

CC peptides (e.g. see AAR72679) were derived from
CC hydrophobicity/flexibility profiles of the published sequence. Peptides
CC containing this sequence can be conjugated to carrier proteins and used
CC to produce antibodies. The antibodies can be used to detect plasma
CC glutathione peroxidase in immunoassays e.g. for detection of selenium
CC deficiency or in diagnosis of certain kidney or liver diseases e.g. graft
CC rejection or certain hepatic cancers.

SO Sequence 226 AA;

Query Match 37.5%; Score 3; DB 16; Length 226;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
|||
Db 71 syxg 74

RESULT 18

AAB22828
ID AAB22828 standard; Protein; 226 AA.

AC AAB22828;

DT 10-JAN-2001 (first entry)

DE Human plasma glutathione peroxidase H (pGPxH).

KW Plasma glutathione peroxidase H; pGPxH; human;
recombinant production.

OS Homo sapiens.

FT Key Location/Qualifiers

FT Misc-difference 73 /label= unknown

FT FT /note= "Encoded by TGA"

PN CN1256312-A.

PD 14-JUN-2000.

PF 29-OCT-1998; 98CN-0121973.

PR 29-OCT-1998; 98CN-0121973.

PA (UYFU-) UNIV FUDAN.

PI Yu L, Tu Q, Fu Q;

DR WPI; 2000-533631/49.

DR N-PSDB; AAA90423.

PT New human glutathione peroxidase and its code sequence, preparation and
PT use -

PS Claim 2; Page 20-21; 26pp; Chinese.

CC This sequence represents human plasma glutathione peroxidase H
CC (pGPxH). The invention relates to this novel human glutathione
CC peroxidase, nucleic acid encoding it, and to recombinant production
CC of human pGPxH. The invention also encompasses applications for
CC human pGPxH.

SO Sequence 226 AA;

Query Match 37.5%; Score 3; DB 21; Length 226;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```
ID AAB08730 standard; Protein: 329 AA.
XX
AC AAB08730;
XX
DT 02-JAN-2001 (first entry)
XX
DE Amino acid sequence of a consensus B7RP1 polypeptide.
XX
KW CRP1; CD28 related protein-1; B7RP1; B7 related protein-1;
KM T-lymphocyte activation; type I transmembrane protein; T cell activation;
KW T cell proliferation; T-cell mediated disorder.
XX
OS Synthetic.
OS Mus sp.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 1..329
FT /note= "Xaa represent unspecified amino acids"
XX
PN W0200046240-A2.
PD 10-AUG-2000.
XX
PF 27-JAN-2000; 2000MO-US01871.
XX
PR 03-FEB-1999; 99US-0244448.
PR 08-MAR-1999; 99US-0264527.
XX
PA (AMGE-) AMGEN INC.
XX
PI Yoshinaga SK;
XX
DR WPI; 2000-543476/49.
XX
PT Novel nucleic acids encoding the proteins CRP-1 and B7RP1 are useful
PT in the treatment, prevention and diagnosis of T cell mediated disorders
PT
PT
XX
PS Disclosure; Page 167-168; 174pp; English.
XX
CC The present sequence represents a consensus B7RP1 (B7 related protein-1)
CC polypeptide. The specification also describes a CRP1 (CD28 related
CC protein-1) polypeptide. The polypeptides are involved in T-lymphocyte
CC activation, and represent a receptor-ligand pair. CRP1 and B7RP1 are
CC predicted to be a type I transmembrane protein. The nucleic acids are
CC useful for regulating T cell activation or proliferation in an animal.
CC The polypeptides are useful for treating, preventing ameliorating or
CC diagnosing a T-cell mediated disorder in an animal. They can also be
CC used to identify test molecules that bind to the polypeptides.
XX
SQ Sequence 329 AA;
Query Match 37.5%; Score 3; DB 21; Length 329;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 syxg 8
| | | |
DB 312 syxg 315
RESULT 22
AAR44893
ID AAR44893 standard; Protein: 533 AA.
XX
AC AAR44893;
XX
DT 22-JUN-1994 (first entry)
XX
DE Diphtheria toxin (delta-147-148; H21X) mutant.
XX
```

```
KW DT; protein exotoxin; NAD-dependent ADP-ribosyltransferase; vaccine;
KW diphtheria toxoid; deletion mutant; muten; variant; double mutant;
KW reversion mutation; site-directed mutagenesis.
XX
OS Corynebacterium diphtheriae.
XX
FH Key Location/Qualifiers
FT Protein 1..533
FT /note= "Diphtheria toxin mutant; Val(147) and
FT Glu(148) have been deleted and His(21)
FT is substcd. by any other amino acid or
FT is absent"
FT
FT Misc-difference 21
FT /note= "any amino acid other than wild-type His
FT or absent"
XX
PN W09325210-A.
PD 23-DEC-1993.
XX
PE 17-MAY-1993; 93WO-US04606.
XX
PR 18-JUN-1992; 92US-0901712.
XX
PA (HARD ) HARVARD COLLEGE.
XX
PI Collier RJ, Killien K, Mekalanos J;
XX
DR WPI; 1994-007178/01.
DR N-PSDB; AA054341.
XX
PT New DNA encoding diphtheria toxin deletion mutants - with no
PT toxicity and low risk of reversion, and derived toxoids and
PT transformed cells, useful in vaccines
PT
PT
XX
PS Claim 11; 42pp; English.
XX
CC Oligonucleotide-directed mutagenesis of the wild-type diphtheria
CC gene results in deletion of the codons for Val-147 and active site
CC residue Glu-148. The resulting muten is not toxic, making it
CC useful in diphtheria vaccines. The risk of reversion to toxicity
CC is much lower for the 147-148 double mutant than for the prior art
CC 148 single mutant, while its immunogenicity is not impaired. The
CC 147-148 muten opt. has other amino acid residues substcd. or
CC deleted, e.g. wild-type His(21) in fragment A. The specification
CC includes the wild-type DT amino acid sequence (see AAR44888) but does
CC not include any mutant sequences; the wild-type sequence was modified
CC according to the description in the claims to give AAR44893.
XX
SQ Sequence 533 AA;
Query Match 37.5%; Score 3; DB 15; Length 533;
Best Local Similarity 100.0%; Pred. No. 2.1e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 syxg 8
| | | |
DB 19 syxg 22
RESULT 23
AAM17581
ID AAM17581 standard; protein: 684 AA.
XX
AC AAM17581;
XX
DT 01-JUL-1997 (first entry)
XX
DE Thermoaerobacter CGTase variant beta-cyclodextrin #1.
XX
KW Cyclomaltoextrin glucanotransferase; CGTase; enzyme; Bacillus circulans;
KW thermoaerobacter; starch; cyclomaltoextrin; cyclodextrin; pesticide;
```

KW	transglycosylation; cyclic glucose oligomer; glucose oligosaccharide;
KW	cholesterol removal; food stabiliser; vitamin stabiliser; fungicide;
KW	plastic laminate; biodegradable plastic; mutain.
XX	Synthetic.
XX	
FH	Key
FT	Modified-site
FT	21 /label= V21X
FT	/note= "X= Phe, Tyr"
FT	Modified-site
FT	47 /label= KA47X
FT	/note= "X= Gln, Leu, Ala, His, Arg"
FT	Modified-site
FT	87 /label= L87X
FT	/note= "X= Ile, His"
FT	Modified-site
FT	88 /label= P88X
FT	/note= "X= Pro, Asn, Lys, His"
FT	Modified-site
FT	89 /label= D89X
FT	/note= "X= Asp, Gly, Ala, Tyr, Lys, Arg, Pro, Glu, or absent"
FT	Modified-site
FT	90 /label= S90X
FT	/note= "X= Gly, Ala, Ser"
FT	Modified-site
FT	91 /label= T91X
FT	/note= "X= Ala, Val, Ser, Gly, Thr"
FT	Misc-difference
FT	92 /label= F91aX
FT	/note= "X= Ala, Val, Gly, Phe, Tyr, or absent"
FT	Modified-site
FT	93 /label= G92V
FT	Modified-site
FT	94 /label= G93X
FT	/note= "X= Gly, Asn, His, Thr or absent"
FT	Modified-site
FT	95 /label= S94X
FT	/note= "X= Gln, Lys, Arg, Trp, Phe, Asn, Ser, or absent"
FT	Modified-site
FT	99 /label= H98X
FT	/note= "X= Gly, Ala"
FT	Modified-site
FT	102 /label= W101X
FT	/note= "X= Gly, Ala"
FT	Modified-site
FT	136 /label= D135L
FT	Modified-site
FT	141 /label= H140X
FT	/note= "X= Ala, Arg, Asn"
FT	Modified-site
FT	146 /label= S145X
FT	/note= "X= Ala, Glu, Trp, Leu"
FT	Misc-difference
FT	147 /note= "possible insertion of Pro, Ala, Ile, Gln, or Ser"
FT	Modified-site
FT	148 /label= E146X
FT	/note= "X= Pro, Ala, Phe, Gln, Ser, Trp, Ile, Arg, Glu, Lys, Asp, Asn, or absent"
FT	Modified-site
FT	149 /label= T147X
FT	/note= "X= Ala, Leu, Ile, Phe, Trp, Gly, Tyr, Arg, Asp, Thr, or absent"
FT	Modified-site
FT	150 /label= D148X
FT	/note= "X= Gly, Asn, Ala, Gln"
FT	Modified-site
FT	151 /label= P149W
FT	Modified-site
FT	152 /label= T150X
FT	/note= "X= Ala, Ser"
FT	Modified-site
FT	169

FT	/label= Y167X
FT	/note= "X= Ala, Phe, Tyr"
FT	Modified-site
FT	170 /label= T168S
FT	Modified-site
FT	180 /label= Y178N
FT	Modified-site
FT	181 /label= G179X
FT	/note= "X= Gly, Ser, Asn, Asp"
FT	Modified-site
FT	182 /label= G180X
FT	/note= "X= Gly, Ser, Asn, Asp"
FT	Modified-site
FT	185 /label= F183X
FT	/note= "X= Phe, Trp, Tyr, Ala"
FT	Modified-site
FT	187 /label= S185X
FT	/note= "X= Pro, His, Arg, Glu, Asp"
FT	Modified-site
FT	194 /label= R192K
FT	Modified-site
FT	195 /label= N193X
FT	/note= "X= Gly, Ala, Asn"
FT	Modified-site
FT	197 /label= F195X
FT	/note= "X= Leu, Ile, Trp, Tyr, Phe"
FT	Modified-site
FT	198 /label= D196X
FT	/note= "X= Ala, Asp, Asn, Ser"
FT	Modified-site
FT	199 /label= L197X
FT	/note= "X= Asp, Glu, Leu"
FT	Modified-site
FT	234 /label= K232X
FT	/note= "X= Lys, Gln, Leu"
FT	Modified-site
FT	235 /label= H233X
FT	/note= "X= His, Gln, Asn, Ile"
FT	Modified-site
FT	261 /label= Y259X
FT	/note= "X= Phe, Trp, Tyr, Ala"
FT	Modified-site
FT	266 /label= E264Q
FT	Modified-site
FT	328 /label= N326X
FT	/note= "X= Gln, Phe, Leu"
FT	Modified-site
FT	371 /label= G370X
FT	/note= "X= Thr, Asn"
FT	Modified-site
FT	372 /label= D371X
FT	/note= "X= Ala, Asp, Ser, Asn, Gly"
FT	Modified-site
FT	374 /label= Y373X
FT	/note= "X= Asp, Glu, Tyr"
FT	Modified-site
FT	376 /label= R375X
FT	/note= "X= Ala, Pro, Gly, Arg, Lys"
FT	Modified-site
FT	598 /label= L600X
FT	/note= "X= unspecified amino acid"
FT	
XX	W09633267-Al.
PN	
XX	24-OCT-1996.
PD	
XX	22-APR-1996;
PF	96WO-DK00179.
XX	
PR	16-NOV-1995;
PR	95DK-0001281.
PR	21-APR-1995;
PR	95DK-0000477.
XX	17-OCT-1995;
XX	95DK-0001173.
PA	(NOVO) NOVO-NORDISK AS.

FT Modified-site 195 /label= N193X
FT /note= "X= Gly, Ala"
FT Modified-site 197 /label= F195X
FT /note= "X= Leu, Ile, Trp, Phe"
FT Modified-site 198 /label= D196X
FT /note= "X= Ala, Gly, Asn, Ser"
FT Modified-site 199 /label= L197X
FT /note= "X= Asp, Glu"
FT Modified-site 234 /label= K232X
FT /note= "X= Gln, Leu"
FT Modified-site 235 /label= H233X
FT /note= "X= Gln, Asn, Ile"
FT Modified-site 261 /label= Y259X
FT /note= "X= Phe, Trp, Ala"
FT Modified-site 266 /label= E264Q
FT /label= N326X
FT /note= "X= Gln, Phe, Leu"
FT Modified-site 371 /label= G370X
FT /note= "X= Thr, Asn"
FT Modified-site 372 /label= D371X
FT /note= "X= Ala, Ser, Asn, Gly"
FT Modified-site 374 /label= Y373X
FT /note= "X= Asp, Glu, Tyr"
FT Modified-site 376 /label= R375X
FT /note= "X= Ala, Pro, Gly, Lys"
FT Modified-site 598 /label= L600X
FT /note= "X= unspecified amino acid"
PN W09633267-A1.
XX
XX
PD 24-OCT-1996.
XX
XX
PF 22-APR-1996: 96MO-DK00179.
XX
XX
PR 16-NOV-1995: 95DK-0001281.
PR 21-APR-1995: 95DK-0000477.
PR 17-OCT-1995: 95DK-0001173.
XX
XX
PA (NOVO) NOVO-NORDISK AS.
XX
XX
PI Andersen C, Dijkhuizen L, Dijkstra BW, Von Der Osten C;
DR WPI: 1996-485774/48.
XX
XX
PT New variants of cyclo:malto:dextrin glucanotransferase (CGTase) -
PT have altered substrate binding, useful for prodn. of cyclodextrin(s)
PT or linear oligosaccharide(s), opt. formed in situ in e.g. baked
PT goods
XX
XX
PS Claim 34; Page -: 161pp; English.
XX
XX
CC AAM17580-W17591, and AAM17606-W17652 represent mutant versions of the
CC cyclomalto-dextrin glucanotransferase (CGTase) of Thermoaerobacter
CC thermophilus sp. ATCC 53627. (see AAM06772 for wild type
CC sequence). CGTase catalyses the conversion of starch and similar
CC substrates into cyclomalto-dextrins (also known as cyclodextrins) via an
CC intramolecular transglycosylation reaction. Cyclodextrins (CD) are cyclic
CC glucose oligomers with a hydrophobic internal cavity that form inclusion
CC complexes with many small hydrophobic molecules. These CGTase mutants

CC have a modified substrate binding and/or product selectivity, compared
CC to this sequence. The mutants are created using primer mutagenesis to
CC modify the gene encoding this sequence. These sequences have greater
CC product selectivity and/or reduced product inhibition (better yields)
CC than wild-type CGTase. These mutant sequences are used to manufacture the
CC 6 main CD types (comprising 6-11 glucose units), or linear 2-12 glucose
CC oligosaccharides, optionally in situ, e.g. during production of baked
CC goods, to stabilise chemicals during their manufacture and in
CC detergents. CD are known for their usefulness in foods, e.g. as a
CC bread-improving agent, to encapsulate/stabilise/solubilise vitamins,
CC dyes, pharmaceuticals, pesticides or fungicides, to bind/remove
CC lipophilic compounds such as cholesterol (e.g. in egg yolk or butter), in
CC plastic laminates, films etc. and to make biodegradable plastics.
XX
SQ Sequence 684 AA:

Query Match 37.5%; Score 3; DB 17; Length 684;
Best Local Similarity 100.0%; Pred. No. 2.5e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 syxg 8
|||
Db 97 syxg 100

RESULT 25
AAM17589
ID AAM17589 standard; protein; 684 AA.
XX
XX
AC AAM17589;
XX
DT 01-JUL-1997 (first entry)
XX
DE Thermoaerobacter CGTase variant beta-cyclodextrin #3.
XX
XX
KW Cyclomalto-dextrin glucanotransferase; CGTase; enzyme; Bacillus circulans;
KW thermoaerobacter; starch; cyclomalto-dextrin; cyclodextrin; pesticide;
KW transglycosylation; cyclic glucose oligomer; glucose oligosaccharide;
KW cholesterol removal; food stabiliser; vitamin stabiliser; fungicide;
KW plastic laminate; biodegradable plastic; muten.
XX
XX
OS Synthetic.
XX
XX
FH Key Location/Qualifiers
FT Modified-site 21 /label= V21X
FT /note= "X= Phe, Tyr"
FT Modified-site 47 /label= K47X
FT /note= "X= Gln, Leu, Ala, His, Arg"
FT Modified-site 87 /label= L87X
FT /note= "X= Ile, His"
FT Modified-site 88 /label= P88X
FT /note= "X= Asn, Lys, His"
FT Modified-site 89 /label= D89X
FT /note= "X= Gly, Ala, Tyr, Lys, Arg, Pro, Glu, or absent"
FT Modified-site 90 /label= S90X
FT /note= "X= Gly, Ala"
FT Modified-site 91 /label= T91X
FT /note= "X= Ala, Val, Ser, Gly"
FT Misc-difference 92 /label= F91aX
FT /note= "X= Ala, Val, Gly, Tyr, or absent"
FT Modified-site 93 /label= G92V
FT Modified-site 94 /label= G93X

FT	Modified-site	/note= "X= Asn, His, Thr or absent"
FT		95
FT		/label= S94X
FT	Modified-site	/note= "X= Gln, Lys, Arg, Trp, Phe, Asn, or absent"
FT		99
FT		/label= H98X
FT	Modified-site	/note= "X= Gly, Ala"
FT		102
FT		/label= W101X
FT	Modified-site	/note= "X= Gly, Ala"
FT		136
FT		/label= D135L
FT	Modified-site	141
FT		/label= H140X
FT		/note= "X= Ala, Arg, Asn"
FT	Modified-site	146
FT		/label= S145X
FT	Misc-difference	/note= "X= Ala, Glu, Trp, Leu"
FT		147
FT	Modified-site	/note= "insertion of Pro, Ala, Ile, Gln, or Ser"
FT		148
FT		/label= E146X
FT	Modified-site	/note= "X= Pro, Ala, Phe, Gln, Ser, Trp, Ile, Arg, Lys
FT		149
FT		Asp, Asn, or absent"
FT	Modified-site	/label= T147X
FT		/note= "X= Ala, Leu, Ile, Phe, Trp, Gly, Tyr, Arg, Asp,
FT		or absent"
FT	Modified-site	150
FT		/label= D148X
FT	Modified-site	/note= "X= Gly, Asn, Ala, Gln"
FT		151
FT		/label= P149W
FT	Modified-site	152
FT		/label= T150X
FT	Modified-site	/note= "X= Ala, Ser"
FT		169
FT		/label= Y167X
FT	Modified-site	/note= "X= Ala, Phe"
FT		170
FT	Modified-site	/label= T168S
FT		180
FT	Modified-site	/label= Y178N
FT		181
FT	Modified-site	/label= G179X
FT		/note= "X= Ser, Asn, Asp"
FT	Modified-site	182
FT		/label= G180X
FT	Modified-site	/note= "X= Ser, Asn, Asp"
FT		185
FT		/label= F183X
FT	Modified-site	/note= "X= Trp, Tyr, Ala"
FT		187
FT		/label= S185X
FT	Modified-site	/note= "X= Pro, His, Arg, Glu, Asp"
FT		194
FT		/label= R192K
FT	Modified-site	195
FT		/label= N193X
FT	Modified-site	/note= "X= Gly, Ala"
FT		197
FT		/label= F195X
FT	Modified-site	/note= "X= Leu, Ile, Trp, Tyr"
FT		198
FT		/label= D196X
FT	Modified-site	/note= "X= Ala, Gly, Asn, Ser"
FT		199
FT		/label= L197X
FT	Modified-site	/note= "X= Asp, Glu"
FT		234
FT		/label= K232X
FT	Modified-site	/note= "X= Gln, Leu"
FT		235

FT	/label= H233X
FT	/note= "X= Gln, Asn, Ile"
FT	261
FT	/label= Y259X
FT	/note= "X= Phe, Trp, Ala"
FT	266
FT	/label= E264Q
FT	328
FT	/label= N326X
FT	/note= "X= Gln, Phe, Leu"
FT	371
FT	/label= G370X
FT	/note= "X= Thr, Asn"
FT	372
FT	/label= D371X
FT	/note= "X= Ala, Ser, Asn, Gly"
FT	374
FT	/label= Y373X
FT	/note= "X= Asp, Glu"
FT	376
FT	/label= R375X
FT	/note= "X= Ala, Pro, Gly, Lys"
FT	598
FT	/label= L600X
FT	/note= "X= unspecified amino acid"
PN	W09633267-A1.
PN	24-OCT-1996.
PD	22-APR-1996; 96WO-DK00179.
XX	
XX	16-NOV-1995; 95DK-0001281.
PR	21-APR-1995; 95DK-0000477.
PR	17-OCT-1995; 95DK-0001173.
XX	
PA	(NOVO) NOVO-NORDISK AS.
XX	
PI	Andersen C, Dijkhuizen L, Dijkstra BW, Von Der Osten C;
XX	WPI, 1996-485774/48.
XX	
PT	New variants of cyclo:malto:dextrin glucanotransferase (CGTase) -
PT	have altered substrate binding, useful for prodn. of cyclodextrin(s)
PT	or linear oligosaccharide(s), opt. formed in situ in e.g. baked
PT	goods
XX	
PS	Claim 40; Page -: 161pp; English.
XX	
CC	AAW17580-W17591, and AAW17606-W17652 represent mutant versions of the
CC	cyclomaltodextrin glucanotransferase (CGTase) of Thermomicrobacter
CC	thermosulphurigenes sp. ATCC 33627. (see AAW06772 for wild type
CC	sequence). CGTase catalyses the conversion of starch and similar
CC	substrates into cyclomaltodextrins (also known as cyclodextrins) via an
CC	intramolecular transglycosylation reaction. Cyclodextrins (CD) are cyclic
CC	glucose oligomers with a hydrophobic internal cavity that form inclusion
CC	complexes with many small hydrophobic molecules. These CGTase mutants
CC	have a modified substrate binding and/or product selectivity, compared
CC	to this sequence. The mutants are created using primer mutagenesis to
CC	modify the gene encoding this sequence. These sequences have greater
CC	product selectivity and/or reduced product inhibition (better yields)
CC	than wild-type CGTase. These mutant sequences are used to manufacture the
CC	6 main CD types (comprising 6-11 glucose units), or linear 2-12 glucose
CC	oligosaccharides, optionally in situ, e.g. during production of baked
CC	goods, to stabilise chemicals during their manufacture and in
CC	degreasers. CD are known for their usefulness in foods, e.g. as a
CC	bread-improving agent, to encapsulate/stabilise/solubilise vitamins,
CC	dyes, pharmaceuticals, pesticides or fungicides, to bind/remove
CC	lipophilic compounds such as cholesterol (e.g. in egg yolk or butter), in
CC	plastic laminates, films etc. and to make biodegradable plastics.
XX	
SD	Sequence 684 AA:

Query Match 37.5%; Score 3; DB 17; Length 684;
Best Local Similarity 100.0%; Pred. No. 2.5e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 syxg 8
|||
Db 97 syxg 100

Search completed: January 14, 2002, 07:56:31
Job time: 726 sec

